GENETIC DIVERSITY AND POPULATION GENETIC STRUCTURE OF TANNER CRAB *CHIONOECETES BAIRDI* IN ALASKAN WATERS

By

Genevieve M. Johnson, B.S.

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in

Fisheries

University of Alaska Fairbanks

May 2019

APPROVED:

J. Andrés López, Ph.D., Committee Chair

Ginny L. Eckert, Ph.D., Committee Member

Sarah M. Hardy, Ph.D., Committee Member

Milo D. Adkison, Ph.D., Chair

Department of Fisheries

S. Bradley Moran, Ph.D., Dean

College of Fisheries and Ocean Sciences

Michael A. Castellini, Ph.D.,

Dean of the Graduate School

Abstract

Tanner crab (*Chionoecetes bairdi*) is a large-bodied species of crab harvested in commercial, personal use, and subsistence fisheries across Alaska. The commercial fisheries were highly productive until the 1980s, when most stocks faced major declines and were closed to harvest. The recovery success of stocks throughout the state has been variable throughout the subsequent decades, leading managers to question whether there are aspects of the population dynamics that are not accounted for. There is limited information on the genetic population structure of *C. bairdi* in Alaskan waters, which has caused uncertainty about whether established management areas align well with distribution and migration patterns for this species. I applied novel high throughput sequencing methods to measure genetic diversity and investigate the genetic population structure of *C. bairdi* in Alaskan waters. Genomic DNA was isolated from samples collected from Southeast Alaska, Prince William Sound, and the Eastern Bering Sea, both east and west of 166°W longitude, and processed according to a Double-Digest Restriction-Associated DNA Sequencing protocol. The final genotype assembly included 89 individuals that were genotyped at 2,740 independent, neutral single-nucleotide polymorphism (SNP) sites, and contained 3.06% missing data. The average observed heterozygosity across SNP sites within regions was significantly lower than the average heterozygosity expected for populations in Hardy-Weinberg equilibrium. An analysis of molecular variance indicated that genetic variability was mostly found within individuals (90%), 10% of variability was observed between individuals within sampling regions, and no significant amount of variation was detected between sampling regions. Furthermore, pairwise F_{ST} estimates between sampling regions were low, and thus the null model of panmixia could not be rejected. Principal components analysis was also congruent with a model of no differentiation among regions. Bayesian analysis implemented in the program STRUCTURE did not support any population partitioning above K = 1 clusters, again indicating that there is not substantial genetic differentiation among the regions sampled from across the state of Alaska. These results indicate high gene flow throughout the distribution of Tanner crab across the Alaska continental shelf. Recognized stocks are genetically indistinguishable from each other. This may indicate that stocks exchange a substantial number of migrants, and may not operate independently. This new information can provide insights as management plans are evaluated and refined.

Table of Contents

List of Figures

Page

List of Tables

Page

Acknowledgements

I am thankful to my advisor, Andrés López, for providing continued guidance throughout this project. He was instrumental in keeping me on task, encouraging me when I encountered difficulties, as well as helping me build a thorough understanding of genomic tools and resources. I would like to thank my committee member, Sarah Hardy, for coordinating the collection of samples for this project, and the National Oceanic and Atmospheric Administration (NOAA) National Marine Fisheries Service (NMFS) and the Alaska Department of Fish and Game (ADF&G) for assisting with sample collection. I am also grateful to my committee member, Ginny Eckert, for providing feedback on the project at various stages, and for pointing me in the direction of relevant crab biology literature.

My fellow students in the López lab made reading genetics literature and working at the lab bench a great joy. I am thankful that I had the opportunity to learn alongside them. My family has provided continuous support through phone calls from thousands of miles away, and I am grateful for their encouragement for me to pursue my academic interests.

This study was funded in part by the U.S. Department of the Interior, Bureau of Ocean Energy Management (BOEM) under cooperative agreements M16AC00004, M16AC00007, and M16AC00011, between BOEM, Alaska Outer Continental Shelf Region, and the University of Alaska Fairbanks. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Government. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Government. This research was supported in part by a UAF Global Change Student Research Grant award with funds from the Cooperative Institute for Alaska Research. The project also received funding from the Northern Gulf of Alaska Applied Research Award, by an anonymous donor. I am grateful for support from the Biomedical Learning and Student Training program through a Graduate Mentoring Research Assistantship. Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Numbers UL1GM118991, TL4GM118992, or RL5GM118990. I would like to thank the Alaska IDeA Network of Biomedical Research Excellence for hosting the 2017 Alaska INBRE Genomics Workshop and providing travel

funding for me to attend. Research reported in this publication was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103395. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the NIH.

General Introduction

Molecular genetic applications have been utilized as a tool for fisheries research for decades (Carvalho & Hauser, 1995; Ovenden et al., 2015). This study continues this line of research by applying novel high-throughput sequencing techniques to the study of population genetic structure of Tanner crab (*Chionoecetes bairdi*) in Alaskan waters. *Chionoecetes bairdi* are a species targeted in Alaskan commercial, subsistence, and personal use fisheries. Their abundances over recent decades have been unstable in certain management regions. It is imperative to improve baseline knowledge of the population structure in this species so that decisions are made with the best possible information.

Marine populations and connectivity

An ongoing focus in marine biology and fisheries research is understanding the spatial and genetic structure of natural populations. Information on population structure is valuable for making decisions regarding the effective management or conservation of a species (Palumbi, 2003; Reiss et al., 2009; Spies & Punt, 2015; Waples et al., 2008). Due to lack of observable physical boundaries and obstacles to migration in marine environments, it was long assumed that marine populations were largely homogenous and panmictic (Caley et al., 1996; Cowen et al., 2000). However, while marine populations often have less marked levels of variation than freshwater species, which have obvious barriers to population connectivity, low-level variation and evidence of subpopulations has been observed in marine organisms (Hutchinson et al., 2001; Merkouris et al., 1998; Reiss et al., 2009; Ward et al., 1994). Even minor levels of differentiation can indicate a lack of connectivity among regions (Knutsen et al., 2011; Palumbi, 2003; Reiss et al., 2009).

For many benthic species that undergo pelagic larval periods, the scale of adult migration is much smaller than the dispersal range of the pelagic larvae (Caley et al., 1996; Cowen & Sponaugle, 2009). This is likely the case for *C. bairdi* as well. While adults move an average 23-40m per day, and are able to move in and out of some bays in Southeast Alaska, habitat barriers such as hard substrate, deep gullies, and shallow ridges could limit exchange of adults between different regions (Taggart et al., 2008). Meanwhile, larval dispersal of *C.bairdi* is

reported to extend over spatial scales that are orders of magnitude larger than any observed adult migration, evidenced by *C. bairdi* larvae identified in plankton tow samples from the Chukchi Sea, 500 km north of the most northern extent of the adult range (Landeira et al., 2018). Thus, quantification and mapping of larval dispersal has been a major focus in the estimation of population structure and connectivity of benthic marine species (Cowen & Sponaugle, 2009). Appropriate modeling approaches for estimating larval dispersal distances are debated (Cowen et al., 2000; Treml et al., 2012). A strong correlation between pelagic larval duration (PLD) and dispersal distance (Shanks et al., 2003) makes intuitive sense. If an individual spends more time in its pelagic larval life stage and is a passive particle, it would be transported by currents for a longer period of time, and travel farther than an individual with a shorter PLD. PLD has thus been incorporated as a parameter into models of population connectivity as a substitute for direct observation of larval dispersal distance, which is difficult to quantify in situ (Lester & Ruttenberg, 2005; Mitarai et al., 2008).

For some species, PLD and larval dispersal are poorly correlated (Cowen & Sponaugle, 2009; Shanks, 2009; Weersing & Toonen, 2009). This lack of relationship may be due to differences in biology among species, such as the swimming behavior of larvae, or due to effects of physical oceanographic factors that trap larvae close to their origin. Additionally, interactions between these terms can also cause PLD to be a poor indicator of actual larval dispersal (Cowen & Sponaugle, 2009; D'Aloia et al., 2015; Shanks, 2009). For example, for a given species, larvae may demonstrate diurnal swimming pattern in some tidal phases, but not others (Rodríguez et al., 1997). Furthermore, PLD for most species, including *C. bairdi,* can vary according to temperature conditions and food availability (Kogane et al., 2005; Urban & Hart, 1999).

Factors influencing larval dispersal of *Chionoecetes bairdi*

Chionoecetes bairdi life history includes four larval life stages, some of which vary in duration according to local conditions. The initial stage, prezoeae, lasts approximately 30 minutes, and is not influenced by local biological conditions; the second and third phases, stage 1 and 2 zoeae, each have a duration of approximately 1 month, but this duration is negatively correlated with temperature and food availability; the final stage, megalopae, has approximately a 1 month or longer duration (Haynes, 1973; Haynes, 1981; Incze et al., 1982; Urban & Hart, 1999). Several

physical mechanisms, including tidal currents, water stratification, shore topography, wind, and other oceanographic features may affect larval dispersal and retention in general (Cowen & Sponaugle, 2009; Parada et al., 2010; Weslawski et al., 2000), and in *C. bairdi* larvae in particular (Bunch et al., 1998; Richar et al., 2015). A model of larval advection patterns of *C. bairdi* in the Eastern Bering Sea (EBS), based on the Regional Ocean Modeling System (ROMS), suggests that larvae are transported north across the continental shelf in the northwest region of the EBS, but have higher retention and isolation in the southeast EBS near Bristol Bay, (Richar et al., 2015). The model does not account for larval behavior which may also influence how the advection patterns through its interaction with the currents.

In addition to physical oceanographic factors, biological constraints may limit successful larval development. Feeding larvae require food input to meet energy requirements for development. Larval *C. bairdi* feed primarily on copepods and other zooplankton, and require a prey density >20 L^{-1} to successfully feed (Incze & Paul, 1983; Paul et al., 1979). When zooplankton concentrations are suboptimal, zoeae can consume large-cell phytoplankton, but do not consume these at a rate fast enough to meet energy demands (Incze & Paul, 1983). The zooplankton assemblage of the EBS is affected by the timing of ice retreat, with early retreats characterizing assemblages favoring small copepods rather than the large, lipid-rich copepods and euphausiids that occur in cool years with a late ice retreat (Coyle et al., 2011; Hunt et al., 2011). *Chionoecetes bairdi* in the EBS could potentially be affected by climate-mediated variability in the types of prey available or the distribution and timing of blooms (Sigler et al., 2016). Correlation analyses do indicate at least some bottom-up climate-driven control in the variability *C. bairdi* recruitment success when environmental parameters are time-lagged to account for to pressures on larval life stages (Zheng & Kruse, 2006).

Settlement and reproduction of *Chionoecetes bairdi*

Population connectivity may be affected by highly variable survival rates at the megalopa stage, because juveniles may settle in unsuitable habitat (Hovel, 2003). Juveniles molt six times during the first year, and molts become less frequent over time (Donaldson et al., 1981). Population structure may also be heavily impacted by mortality or shifts in occupied habitat due to predation at juvenile stages. For example, in 1981, 1984, and 1985, an estimated 84%, 95%, and 94% of

age 1 juvenile *C. bairdi* in regions of the Bering Sea were lost due to predation by Pacific cod (*Gadus macrocephalus*) (Livingston, 1989; Urban & Hart, 1999). Predation on juveniles by adult *C. bairdi* is also known to occur, but juveniles in Southeast Alaska tend to occupy different depths than adult females, possibly as a means to avoid conspecific predation (Nielsen et al., 2007).

Age can only be estimated, because no hard parts are retained between molts. Estimates for lifespan and age at maturity based on regressions of growth rate extrapolated to observed adult carapace sizes suggests that males could live for 12 years or more, and females could live at least 9 years (Donaldson et al., 1981; Urban & Hart, 1999). Upon reaching maturity, females have a terminal molt during which the abdominal flap widens. This is known as the primiparous stage. Mating usually occurs during this terminal molt. Males and primiparous females mate in pairs. The male clasps the female for several hours and sometimes assists with the female molt. Mating happens when the female is in soft shell condition. Females continue to produce one egg clutch per year, and are then referred to as multiparous females. Females off the coast of Kodiak aggregate in mounds of hundreds of individuals to release larvae and to mate (Stevens et al., 1994). Mating between males and multiparous females is aggressive, and the females often resist (Donaldson & Adams, 1989). Interestingly, multiparous females can retain and store sperm from a previous mating event for up to three years, but fecundity is reduced (Paul, 1984). This behavior can have effects on population genetic estimates that commonly rely on the assumption of random mating. Typical annual fecundity for females is between 24,000 - 318,000 embryos, which develop on the female's abdomen for a year until the next mating season (Urban & Hart, 1999). The timing of these spawning events and the release of larvae have been linked to tidal current patterns in the Kodiak area, which suggests larval dispersal could be optimized for transport by currents (Stevens, 2003). There is genetic evidence from mitochondrial haplotypes that these point-source releases of larvae into the Alaska Coastal Current (ACC) could be a major driver of genetic population structure. Haplotype diversity matches a pattern of unidirectional flow that would be expected if reproductively successful migrants were transported in an eastern to western direction along the coast of Alaska (Bunch et al., 1998; Park et al., 2007).

Genetic structure of marine populations

Estimates of larval dispersal distance based on PLD alone are unreliable, and direct observation of larval dispersal is logistically challenging, if not impossible, but natural markers provide an indirect way of measuring fluxes of individuals between different regions. These natural tags include isotopes, DNA, and otoliths (Cowen & Sponaugle, 2009). Crab do not have hard structures, such as otoliths, that remain with an individual through its life, but DNA markers do provide a useful tool to investigate population structure. Variable sites across the genome can be identified and characterized to compare aggregations of a species, and identify genetic heterogeneity that may be the product of intra-specific barriers to movement of reproductively successful individuals (Barber et al., 2002; Knutsen et al., 2011; Xuereb et al., 2018).

The use of genetic data to infer elements of population structure is well established. The earliest theoretical methods, which are still relevant, were developed by Sewall Wright. Wright (1951) introduced a set of test statistics to quantify genetic similarities and differences at different population scales, which he termed F-statistics. They are broken into F_{IS} (the homozygosity due to inbreeding within a subpopulation), F_{ST} (the similarity of alleles within a population compared to the total population), and F_{IT} (the homozygosity in the total population due to inbreeding). F_{ST} has become the most ubiquitous statistic used to describe population structure. Wright's model of metapopulations is referred to as the island model, where each subpopulation represents an "island", and migrants are exchanged at the same rate between all populations. His formula for F_{ST} also includes an estimate of the exchange of breeding migrants between subpopulations, N_m (Wright, 1951). Weir and Cockerham (1984) developed a method to estimate Wright's F_{ST} given multilocus allele frequencies from subsampled populations. Parameters such as linkage disequilibrium (non-independent loci) can be used to estimate the effective population size (minimum size of a population under Hardy-Weinberg assumptions that would result in the observed genetic variability) of subpopulations. There are now several other models and expected patterns of metapopulation structure. Some represent population structures as discrete types. For instance, Hellberg et al. (2002) described six different patterns of genetic differentiation based on migration rates and effective population size. Other approaches represent population discreteness along a continuum (Waples & Gaggiotti, 2006). The isolation-by-distance (IBD) model represents populations in this way, so it is useful for characterizing

many real-world situations (Cunningham et al., 2009). However, because IBD population structure is modeled along a continuum, defining boundaries based on these models is not straightforward (Spies et al., 2015).

Over recent decades, microsatellites were the most widely preferred marker for studies on population structure because they are functionally neutral and can have high rates of polymorphism (DeFaveri et al., 2013). Microsatellite markers must be identified in the genome of an unexplored species before they can be applied to an analysis, and for this reason, most microsatellite studies of wild populations include a relatively small set of loci (Koskinen et al., 2004). Initially, microsatellites were considered for use as genetic markers in this study. However, only two microsatellite loci have been successfully amplified in *C. bairdi* (Puebla et al., 2003). Ideally, at least ten additional loci would need to be identified and characterized to support a robust study of population genetics (Koskinen et al., 2004). However, the type of genetic marker used significantly affects the outcome of estimates of marine population connectivity. Microsatellites consistently yielded lower estimates of genetic differentiation between regions than mitochondrial sequence markers (Weersing & Toonen, 2009). So, even a larger suite of microsatellite loci may not be sensitive enough to detect small differences in gene diversity that are characteristic of species with potential for long-distance larval dispersal. Rather than spending resources to develop an extensive set of species-specific primers, we turn to new applications of high throughput sequencing.

Reduced-representation sequencing approaches use modern sequencing platforms to conduct genome-wide scans for simultaneous discovery and genotyping of hundreds to thousands of single nucleotide polymorphisms (SNPs). These approaches are particularly useful for resolving fine scale structure, or for detecting low levels of variation (Hemmer-Hansen et al., 2014; Peterson et al., 2012; Rodríguez-Ezpeleta et al., 2016). Double digest restriction associated DNA sequencing (ddRADseq) is a type of reduced-representation sequencing, meaning that, rather than sequencing all regions of the genome, the method targets a subset of regions across the genome. ddRADseq provides a cost-effective way to analyze the genomes of non-model organisms with little reference data, because consistent sequence reads can be obtained from genome regions that are adjacent to the restriction enzyme cut sites. The restriction enzyme

digestion is based on random occurrences of restriction enzyme recognition sites that are relatively consistent throughout the genomes of closely related individuals, so there is no primer-identification step required. Size selection of the digested fragments reduces the proportion of the genome that is sequenced, reducing the resources necessary for large scale studies (Peterson et al., 2012). The combination of these factors has led to ddRADseq quickly becoming the method of choice to examine population genetic structure in wild populations.

Population structure and management

In the Bering Sea and Aleutian Islands, *C. bairdi* are jointly managed by the National Oceanic and Atmospheric Administration's (NOAA) National Marine Fisheries Service (NMFS) and the Alaska Department of Fish and Game (ADF&G), whereas the stocks in the Gulf of Alaska are solely managed by ADF&G. *Chionoecetes bairdi* is the only personal use and subsistence harvested species of the *Chionoecetes* genus in the state of Alaska (Urban & Hart, 1999). The fisheries are targeted for male-only harvest. The commercial harvest in Southeast Alaska was historically much lower than the harvest in the Bering Sea, but it has been more stable over time (Figure 1). The Bering Sea commercial fishery was once one of the largest crab fisheries in the state in terms of landed weight, but it has faced mixed success over the past few decades (NPFMC, 2016; Stockhausen, 2018). Commercial harvests peaked in the 1970s with a maximum harvest of 20,537t (Otto, 1990; Woodby et al., 2005). Stocks throughout the state have undergone fluctuations in abundance. The EBS fishery has been subject to periodic closures, and the Prince William Sound (PWS) fishery has been closed to commercial harvest since 1988 (Rumble et al., 2014; Stockhausen, 2018). The Bering Sea stocks were closed to commercial harvest over the 2016/2017 season. In the 2017/2018 season, targeted commercial harvest was allowed in the area west of 166° W, while harvest remained closed in the area east of 166° W, because mature male biomass estimates did not meet minimum requirements to open the fishery in that area (Westphal & Nichols, 2018). The PWS region opened a test fishery for commercial harvest in 2016 (Russ et al., 2017).

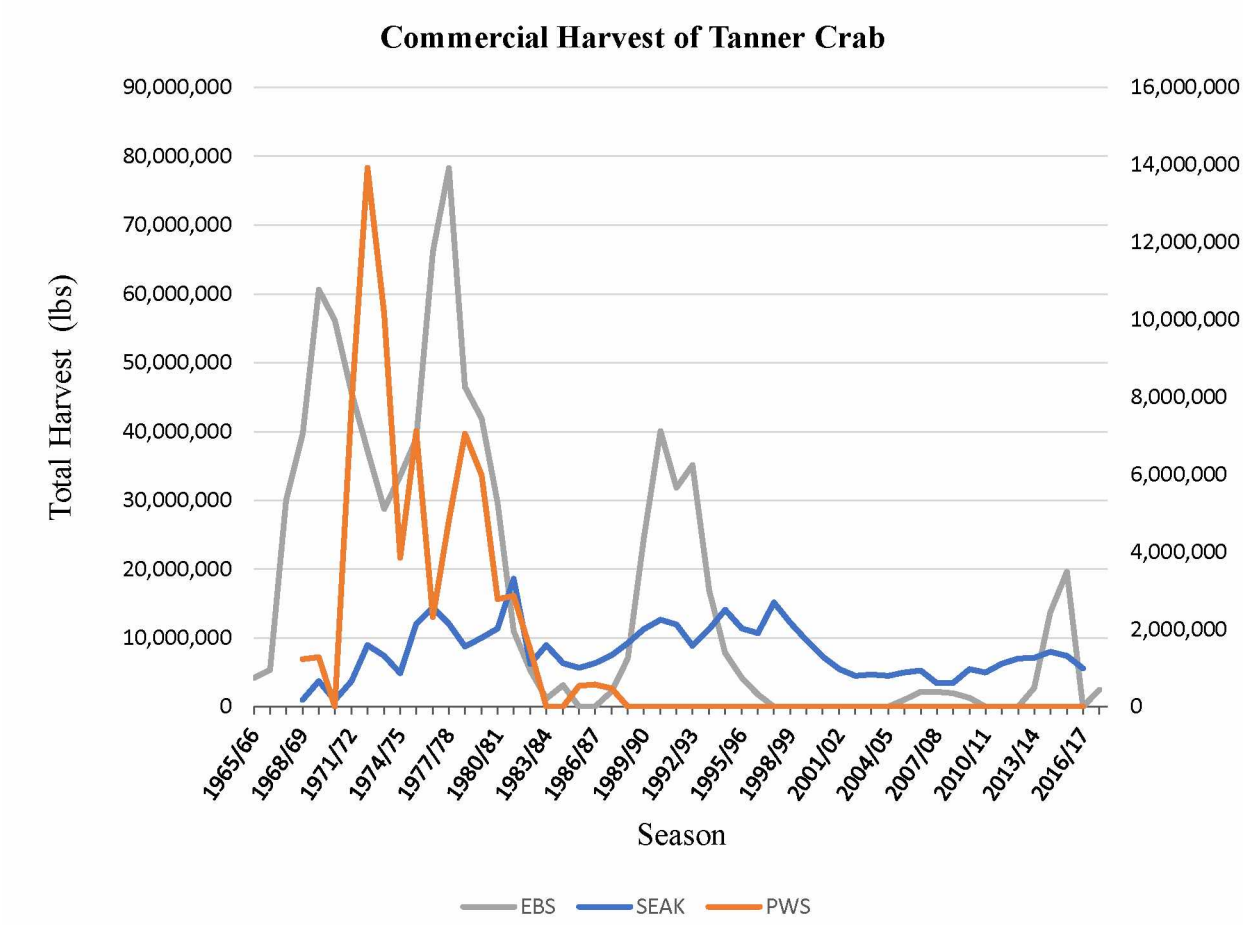

Figure 0.1: Commercial harvest of Tanner crab from sampling regions

Southeast Alaska (SEAK), Prince William Sound (PWS), and the Eastern Bering Sea (EBS; East and West of 166° W combined) districts. SEAK and PWS harvest totals are represented on the scale in the right axis, while EBS harvest is represented on that in the left axis. Harvest records are compiled from Rumble et al., 2014, Stockhausen, 2018, and Wood et al., 2017.

Insight on population genetic structure may be able to uncover multi-generational movement patterns that are not otherwise distinguishable, and may provide indication of why the stock abundances are so variable. Yet, methods for integrating population structure information obtained through genetic analysis to species management are not well-established, and it is important to note that statistically significant differentiation may not equate to biologically significant differentiation (Waples, 1998; Waples et al., 2008). Spies et. al., (2015) performed simulations to test how much weight population connectivity should be given when forming management plans for species that demonstrate isolation by distance. The models suggested that when there is even weak evidence for distinction, it is better to manage those groups separately.

It is possible to manage stocks along boundaries that do not match biological breaks, but often over-splitting stocks will require a larger investment of resources than necessary, while under-splitting stocks could lead to an unintended/unknown overexploitation of a distinct subpopulation that is lumped into a larger population (Reiss et al, 2009; Spies & Punt., 2015; Waples, 1998). In the worst cases, simulations show that undetected subpopulations could go extinct and be masked by the appearance of a minor decline in the overall stock. The simulations also revealed that the loss of subpopulations could have impact on the overall stock's ability to recover from a collapse, because of substantial recruitment declines (Frank & Brickman, 2000).

Mismatches between management areas and actual population structure can arise because fisheries managers must consider economic and political factors in addition to biological characteristics. Sometimes the criteria used to determine biological structure for the purpose of designating stocks are not suitable for the overarching management goals (Reiss et al., 2009). Waples & Gaggiotti (2006) brought attention to the disparity in definitions of the term "population" throughout scientific literature. The ecological definition of a population is based on co-occurrence and interaction of individuals in space and time, whereas the evolutionary definition is based on genetic structure and reproductive interactions of individuals. These can be studied by different methods, yielding potentially conflicting results. The tools commonly used by fisheries biologists to assess populations often measure ecologically relevant population structure. However, Reiss et al. (2009) argue that evolutionary population structure could be more relevant and better linked to common goals of management, long-term sustainability of the stock, and year-to-year stability of the spawning biomass. In simulations, designing management boundaries that accounted for known genetic population structure led to lower probabilities of overharvesting, improved likelihood of recovery of previously masked subpopulations, and higher harvest overall (Spies & Punt, 2015). The ADF&G 2011 regional long-term Alaska crab research priorities, included specific goals to use population and landscape genetics information to improve delineation of stock structure and variation in Tanner crab (Webb & Woodby, 2011).

Evaluating the genetic population structure of *C. bairdi* is important because in some regions, there may be more discrete structure than is currently assumed, or alternatively there could be

sources of juvenile recruitment to a region that are inadvertently ignored. Red king crab (*Paralithodes camtschaticus*), which occupy many of the same areas as *C. bairdi* in Southeast Alaska, have fine-scale population structure within the region. There is evidence of almost no exchange of migrants in or out of certain bays, likely because of oceanographic patterns within those bays that limit larval dispersal (Grant & Cheng, 2012). Bunch et al. (1998) found that *C. bairdi* haplotype diversity was correlated with directional movement by the ACC. As an example, the most dominant haplotype in Southeast Alaska represented 85% of total haplotype diversity in that region. That haplotype was present at lower levels in more western sampling regions: 43% at Kodiak, 25% at Cook Inlet, and 22% in Bristol Bay. Assuming an island model, this pattern suggests unidirectional flow of migrants aligned with the ACC flow. Perhaps the most interesting region in which to apply population genomics to management questions is the Bering Sea. Overfishing limits are set for the EBS based on NMFS surveys of the whole area, and total allowable catch is then set for two subunits, east and west of 166°W. Evidence from three allozyme markers demonstrated significant differentiation between these regions (Merkouris et al., 1998). The boundary at 166°W was initially set because of differences in size-at-maturity on either side (Somerton, 1980; Somerton, 1981), but that life-history variation appears to occur along a gradient, and size-at-maturity also changes over time (Otto & Pengilly, 2002; Zheng, 2008). There is currently a lack of consensus about whether the boundary between subunits reflects actual biological distinctions, or whether the whole EBS should be managed as one stock (NPFMC, 2016; Webb & Woodby, 2011). Additionally, *C. bairdi* are known to hybridize with *C. opilio* on both sides of the boundary (NPFMC, 2016), and microsatellite analysis of *C. opilio* indicated that the population is panmictic throughout the Bering, Chukchi, and Beaufort Seas (Albrecht et al., 2014). Considering these factors, it is reasonable to expect *C. bairdi* to display panmictic structure or low levels of genetic differentiation across the Bering Sea as well.

Hybridization with *Chionoecetes opilio*

Snow crab (*Chionoecetes opilio*) range from the Bering Sea through the Arctic and Atlantic oceans. In the southern part of their range in the Bering Sea, they overlap with *C. bairdi* and are known to hybridize (Urban et al., 2002). Bentzen and Jensen (1996) found that a greater proportion of *opilio x bairdi* hybrids carried *C. opilio* mitochondrial haplotypes. The

mitochondrial genome is maternally inherited, so this finding implies that most interspecific pairings occur between a male *C. bairdi* and a female *C. opilio*. This is reasonable because of the larger size of *C. bairdi* relative to *C. opilio* and aggressive mating strategy of *Chionoecetes* (Donaldson & Adams, 1989; Urban & Hart, 1999). The larger *C. bairdi* males may be able to compete well and have control when clasping the smaller *C. opilio* females, whereas small *C. opilio* males may be less successful at clasping larger *C. bairdi* females. Allozymes across *Chionoecetes* populations show evidence of introgression and hybrid backcrossed individuals (Merkouris et al., 1998). Thus, the hybrids are viable and contribute to both *C. opilio* and *C. bairdi* populations. The extent of this contribution to populations of both species is not yet determined.

Hybrids, identified based eye stalk color and morphometric characteristics of the carapace, are sorted and counted during annual surveys. In the most recent annual survey, hybrids were captured at 102 of the 375 sampling stations (NPFMC, 2016). The Alaska Department of Fish and Game produced a guide to identify hybrids in the field based on carapace morphology and features. The physical characteristics used to identify hybrids tend to be intermediate expressions of *C. bairdi* and *C. opilio* phenotypes. However, when accuracy of the method was tested, there was a high variance in accuracy of experts and trainees (Urban et al., 2002). The phenotypic characteristics of putative hybrids occur across a spectrum. They can be subtle, because hybrids are reproductively viable and can backcross with *C. opilio* and *C. bairdi* (Merkouris et al., 1998; Urban et al., 2002). Thus, survey estimates of the prevalence of hybridization could be underestimating its true magnitude across these populations, especially F_2 and later generation hybrids. To complicate matters, hybrids are considered legal *C. opilio* catch, and since crossed individuals are generally larger than *C. opilio,* there may be incentive to retain them with the catch (Smith et al., 2005). The present study generated data that can be used to improve knowledge of population dynamics and interactions between these species.

Reasons for a population genomics study

Crab fisheries are an important part of Alaska's culture. Commercial crab fisheries make significant contributions to the local economies. Unfortunately, Tanner crab stocks have faced repeated harvest closures since the 1980s because minimum biomass indices are not consistently

met. Fluctuations in year-to-year abundance leads some fishery managers to question whether the identified stocks match biologically significant population segments within the species distribution (NPFMC, 2016; Webb & Woodby, 2011). Ideally, fisheries management strategies align with boundaries of stocks that are independent and respond independently to harvest pressures (Carvalho & Hauser, 1995). Where populations are freely mixing, management should instead be able to account the degree and extent of movement of individuals between regions when designating harvest strategies (Ovenden et al., 2015). The analysis of genome-wide variation within *C. bairdi* presented here provides the most detailed characterization of genetic population structure of this species to date.

References

ADF&G. (Retrieved 2019, Feb 1). Southeast Alaska Tanner Crab Harvest – Registration Area A (Southeast Alaska). Retrieved from https://www.adfg.alaska.gov/index.cfm?adfg=commercialbyareasoutheast.shellfish_harvest_tannera

Albrecht, G. T., Valentin, A. E., Hundertmark, K. J., & Hardy, S. M. (2014). Panmixia in Alaskan populations of the snow crab Chionoecetes opilio (Malacostraca: Decapoda) in the Bering, Chukchi, and Beaufort Seas. *Journal of Crustacean Biology*, *34*(1), 31-39.

Barber, P. H., Palumbi, S. R., Erdmann, M. V., & Moosa, M. K. (2002). Sharp genetic breaks among populations of Haptosquilla pulchella (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology*, *11*(4), 659-674.

Bentzen, P., & Jensen, P. (1996). Development of genetic markers for crabs. Final Report to Alaska Department of Fish and Game Genetics Lab.

Bunch, T., Highsmith, R. C., & Shields, G. F. (1998). Genetic evidence for dispersal of larvae of Tanner crabs (Chionoecetes bairdi) by the Alaskan Coastal Current. *Molecular Marine Biology and Biotechnology*, 7(2), 153-159.

Caley, M. J., Carr, M. H., Hixon, M. A., Hughes, T. P., Jones, G. P., & Menge, B. A. (1996). Recruitment and the local dynamics of open marine populations. *Annual Review of Ecology and Systematics*, 477-500.

Carvalho, G. R., & Hauser, L. (1995). Molecular genetics and the stock concept in fisheries. In *Molecular genetics in fisheries* (pp. 55-79). Springer, Dordrecht.

Cowen, R. K., Lwiza, K. M., Sponaugle, S., Paris, C. B., & Olson, D. B. (2000). Connectivity of marine populations: open or closed?. *Science*, *287*(5454), 857-859.

Cowen, R. K., & Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Marine Science*, *1*.

Coyle, K. O., Eisner, L. B., Mueter, F. J., Pinchuk, A. I., Janout, M. A., Cieciel, K. D., & Andrews, A. G. (2011). Climate change in the southeastern Bering Sea: impacts on pollock stocks and implications for the oscillating control hypothesis. *Fisheries Oceanography*, *20*(2), 139-156.

Cunningham, K. M., Canino, M. F., Spies, I. B., & Hauser, L. (2009). Genetic isolation by distance and localized fjord population structure in Pacific cod (Gadus macrocephalus): limited effective dispersal in the northeastern Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(1), 153-166.

D'Aloia, C. C., Bogdanowicz, S. M., Francis, R. K., Majoris, J. E., Harrison, R. G., & Buston, P. M. (2015). Patterns, causes, and consequences of marine larval dispersal. *Proceedings of the National Academy of Sciences*, *112*(45), 13940-13945.

DeFaveri, J., Viitaniemi, H., Leder, E., & Merilä, J. (2013). Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. *Molecular ecology resources*, *13*(3), 377-392.

Donaldson, W. E., Cooney, R. T., & Hilsinger, J. R. (1981). Growth, age and size at maturity of Tanner crab, Chionoecetes bairdi MJ Rathbun, in the northern Gulf of Alaska (Decapoda, Brachyura). Crustaceana, 40(3), 286-302.

Donaldson, W. E., & Adams, A. E. (1989). Ethogram of behavior with emphasis on mating for the Tanner crab Chionoecetes bairdi Rathbun. *Journal of Crustacean Biology*, 37-53.

Frank, K. T., & Brickman, D. (2000). Allee effects and compensatory population dynamics within a stock complex. Canadian Journal of Fisheries and Aquatic Sciences, 57(3), 513-517.

Grant, W. S., & Cheng, W. (2012). Incorporating deep and shallow components of genetic structure into the management of Alaskan red king crab. *Evolutionary applications*, *5*(8), 820-837.

Haynes, E. (1973). Description of the prezoea and stage I zoea of Chionoecetes bairdi, and C. opilio (Oxyrhyncha, Oregoniinae). *Fishery Bulletin*, *71*(3), 769-775.

Haynes, E. (1981). Description of the stage II zoea of snow crab, Chionoecetes bairdi, (Oxyrhrncha, Majidae), from plankton of lower Cook Inlet, Alaska. *Fishery Bulletin*, *79*(1), 177-182.

Hellberg, M. E., Burton, R. S., Neigel, J. E., & Palumbi, S. R. (2002). Genetic assessment of connectivity among marine populations. *Bulletin of marine science*, *70*(Supplement 1), 273-290.

Hemmer-Hansen, J., Therkildsen, N. O., & Pujolar, J. M. (2014). Population genomics of marine fishes: next-generation prospects and challenges. *The Biological Bulletin*, *227*(2), 117-132.

Hovel, K. A. (2003). Habitat fragmentation in marine landscapes: relative effects of habitat cover and configuration on juvenile crab survival in California and North Carolina seagrass beds. *Biological Conservation*, *110*(3), 401-412.

Hunt, G. L., Coyle, K. O., Eisner, L. B., Farley, E. V., Heintz, R. A., Mueter, F., & Stabeno, P. J. (2011). Climate impacts on eastern Bering Sea foodwebs: a synthesis of new data and an assessment of the Oscillating Control Hypothesis. *ICES Journal of Marine Science: Journal du Conseil*, fsr036.

Hutchinson, W. F., Carvalho, G. R., & Rogers, S. I. (2001). Marked genetic structuring in localised spawning populations of cod Gadus morhua in the North Sea and adjoining waters, as revealed by microsatellites. *Marine Ecology Progress Series*, *223*, 251-260.

Incze, L. S., Armstrong, D. A., & Wencker, D. L. (1982). Rates of development and growth of larvae of Chionoecetes bairdi and C. opilio in the southeastern Bering Sea. In *Proc. Int. Symp. Genus Chionoecetes. Fairbanks, AK: U. of Alaska Sea Grant College Program Rep* (Vol. 82, No. 10, pp. 191-218).

Incze, L. S., & Paul, A. J. (1983). Grazing and predation as related to energy needs of stage I zoeae of the Tanner crab Chionoecetes bairdi (Brachyura, Majidae). *The Biological Bulletin*, *165*(1), 197-208.

Knutsen, H., Olsen, E. M., Jorde, P. E., Espeland, S. H., André, C., & Stenseth, N. C. (2011). Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. *Molecular ecology*, *20*(4), 768-783.

Kogane, T., Hamasaki, K., & Nogami, K. (2005). Effect of temperature on survival and developmental period of larval snow crab Chionoecetes opilio (Brachyura: Majidae) reared in the laboratory. *Nippon Suisan Gakkaishi*, *71*(2), 161-164.

Koskinen, M. T., Hirvonen, H., Landry, P. A., & Primmer, C. R. (2004). The benefits of increasing the number of microsatellites utilized in genetic population studies: an empirical perspective. *Hereditas*, *141*(1), 61-67.

Landeira, J. M., Matsuno, K., Tanaka, Y., & Yamaguchi, A. (2018). First record of the larvae of tanner crab Chionoecetes bairdi in the Chukchi Sea: A future northward expansion in the Arctic?. *Polar Science*, *16*, 86-89.

Livingston, P. A. (1989). Interannual trends in Pacific cod, Gadus macrocephalus, predation on three commercially important crab species in the eastern Bering Sea. *Fishery Bulletin*, *87*(4), 807-827.

Lester, S. E., & Ruttenberg, B. I. (2005). The relationship between pelagic larval duration and range size in tropical reef fishes: a synthetic analysis. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*(1563), 585-591.

Merkouris, S. E., Seeb, L. W., & Murphy, M. C. (1998). Low levels of genetic diversity in highly exploited populations of Alaskan Tanner crabs, Chionoecetes bairdi, and Alaskan and Atlantic snow crabs, C. opilio. *Fishery Bulletin: National Oceanic and Atmospheric Administration, 96*, 525-537.

Mitarai, S., Siegel, D. A., & Winters, K. B. (2008). A numerical study of stochastic larval settlement in the California Current system. *Journal of Marine Systems*, *69*(3-4), 295-309.

Nielsen, J. K., Taggart, S. J., Shirley, T. C., & Mondragon, J. (2007). Spatial distribution of juvenile and adult female Tanner crabs (Chionoecetes bairdi) in a glacial fjord ecosystem: implications for recruitment processes. ICES Journal of Marine Science: Journal du Conseil, 64(9), 1772-1784.

NPFMC The Plan Team for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands, K. Bush, M. Dorn, G. Eckert, R.J. Foy, B. Garber-Yonts, T. Hamazaki, J. N. Ianelli, A.E. Punt, L. Rugolo, M.S.M. Siddeek, W. Stockhausen, L. Slater, D. Stram, C. Szuwalski, B. J. Turnock, D. Weber and J. Zheng. (2016). "2016 Crab SAFE - North Pacific Fishery Management Council." http://www.npfmc.org/wp-content/PDFdocuments/resources/SAFE/CrabSAFE/2016CrabSAFE_final.pdf

Otto, R. S. (1990). An overview of eastern Bering Sea king and Tanner crab fisheries. In *Proceedings of the international symposium on king and Tanner crabs. University of Alaska Fairbanks, Alaska Sea Grant Program Report* (pp. 90-4).

Otto, R. S., & Pengilly, D. (2002). Spatiotemporal trends in Tanner crab (Chionoecetes bairdi) size at maturity. *Crabs in cold water regions: biology, management, and economics. AJ Paul, EG Dawe, R. Elner, GS Jamieson, GH Kruse, RS Otto, B. Sainte-Marie, TC Shirley and D. Woodby (eds.). University of Alaska Sea Grant, AKU-W-01-003, Fairbanks*, 339-349.

Ovenden, J. R., Berry, O., Welch, D. J., Buckworth, R. C., & Dichmont, C. M. (2015). Ocean's eleven: a critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. *Fish and Fisheries*, *16*(1), 125-159.

Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological applications*, S146-S158.

Parada, C., Armstrong, D. A., Ernst, B., Hinckley, S., & Orensanz, J. M. (2010). Spatial dynamics of snow crab (Chionoecetes opilio) in the eastern Bering Sea—putting together the pieces of the puzzle. *Bulletin of Marine Science*, *86*(2), 413-437.

Park, W., Douglas, D. C., & Shirley, T. C. (2007). North to Alaska: evidence for conveyor belt transport of Dungeness crab larvae along the west coast of the United States and Canada. *Limnology and oceanography*, *52*(1), 248-256.

Paul, A. J., Paul, J. M., Shoemaker, P. A., & Feder, H. M. (1979). Prey concentrations and feeding response in laboratory-reared stage-one zoeae of king crab, snow crab, and pink shrimp. *Transactions of the American Fisheries Society*, *108*(5), 440-443.

Paul, A. J. (1984). Mating frequency and viability of stored sperm in the Tanner crab Chionoecetes bairdi (Decapoda, Majidae). *Journal of Crustacean Biology*, *4*(3), 375-381.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS one, 7(5), e37135.

Puebla, O., Parent, E., & Sévigny, J. M. (2003). New microsatellite markers for the snow crab Chionoecetes opilio (Brachyura: Majidae). Molecular Ecology Notes, 3(4), 644-646.

Reiss, H., Hoarau, G., Dickey-Collas, M., & Wolff, W. J. (2009). Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries*, *10*(4), 361-395.

Richar, J. I., Kruse, G. H., Curchitser, E., & Hermann, A. J. (2015). Patterns in connectivity and retention of simulated Tanner crab (Chionoecetes bairdi) larvae in the eastern Bering Sea. *Progress in Oceanography*, *138*, 475-485.

Rodríguez, A., Drake, P., & Arias, A. M. (1997). Reproductive periods and larval abundance patterns of the crabs Panopeus africanus and Uca tangeri in a shallow inlet (SW Spain). *Marine Ecology Progress Series*, *149*, 133-142.

Rodríguez-Ezpeleta, N., Bradbury, I. R., Mendibil, I., Álvarez, P., Cotano, U., & Irigoien, X. (2016). Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: effects of sequence clustering parameters and hierarchical SNP selection. *Molecular ecology resources*.

Rumble, J., Wessel, M. E. L., Russ, E. A., Goldman, K. J., Gustafson, R. L., & Russ, C. (2014). *Cook Inlet and Prince William Sound Area Management Report for Tanner and King Crab Fisheries through 2013*. Alaska Department of Fish and Game, Division of Sport Fish, Research and Technical Services.

Russ, E., Rumble, J., & Russ, C. (2017, March 20) Memorandum: PWS Tanner Crab Test Fishery Summary. *Alaska Department of Fish and Game, Division of Commercial Fisheries, Homer*.

Shanks, A. L., Grantham, B. A., & Carr, M. H. (2003). Propagule dispersal distance and the size and spacing of marine reserves. *Ecological applications*, S159-S169.

Shanks, A. L. (2009). Pelagic larval duration and dispersal distance revisited. *The biological bulletin*, *216*(3), 373-385.

Sigler, M. F., Napp, J. M., Stabeno, P. J., Heintz, R. A., Lomas, M. W., & Hunt, G. L. (2016). Variation in annual production of copepods, euphausiids, and juvenile walleye pollock in the southeastern Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, *134*, 223-234.

Smith, C. T., Grant, W. S., & Seeb, L. W. (2005). A Rapid, High-Throughput Technique for Detecting Tanner Crabs Chionoecetes bairdi Illegally Taken in Alaska's Snow Crab Fishery. *Transactions of the American Fisheries Society*, *134*(3), 620-623.

Somerton, D. A. (1980). A computer technique for estimating the size of sexual maturity in crabs. *Canadian Journal of Fisheries and Aquatic Sciences*, *37*(10), 1488-1494.

Somerton, D. A. (1981). Regional variation in the size of maturity of two species of tanner crab (Chionoecetes bairdi and C. opilio) in the eastern Bering Sea, and its use in defining management subareas. *Canadian Journal of Fisheries and Aquatic Sciences*, *38*(2), 163-174.

Spies, I., & Punt, A. E. (2015). The utility of genetics in marine fisheries management: a simulation study based on Pacific cod off Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(9), 1415-1432.

Spies, I., Spencer, P. D., & Punt, A. E. (2015). Where do we draw the line? A simulation approach for evaluating management of marine fish stocks with isolation-by-distance stock structure. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(7), 968-982.

Stevens, B. G., Haaga, J. A., & Donaldson, W. E. (1994). Aggregative mating of Tanner crabs, Chionoecetes bairdi. *Canadian Journal of Fisheries and Aquatic Sciences*, *51*(6), 1273-1280.

Stevens, B. G. (2003). Timing of aggregation and larval release by Tanner crabs, Chionoecetes bairdi, in relation to tidal current patterns. *Fisheries Research*, *65*(1), 201-216.

Stockhausen, W. T. (2018). "2018 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions" BSAI Crab SAFE https://www.npfmc.org/wp-content/PDFdocuments/resources/SAFE/CrabSAFE/2018/3-Tanner_SAFE_2018.pdf

Taggart, S. J., Mondragon, J., Andrews, A. G., & Nielsen, J. K. (2008). Spatial patterns and movements of red king and Tanner crabs: implications for the design of marine protected areas. *Marine Ecology Progress Series*, *365*, 151-163.

Treml, E. A., Roberts, J. J., Chao, Y., Halpin, P. N., Possingham, H. P., & Riginos, C. (2012). Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations.

Urban, D., & Hart, D. (1999). Biology of the Tanner Crab, Chionoecetes Bairdi, in Alaska: A Report to the Alaska Board of Fisheries. Alaska Department of Fish and Game, Division of Commercial Fisheries.

Urban, D., Pengilly, D., Jadamec, L., & Byersdorfer, S. C. (2002). Testing carapace morphology characteristics for field identification of Chionoecetes hybrids. In A. J. Paul, E. G. Dawe, R. Elner, G. S. Jamieson, G. H. Kruse, R. S. Otto, & B. Sainte-Marie (Eds.), Crabs in coldwater regions: biology, management and economics. Lowell Wakefield Fisheries Symposium Series. Alaska Sea Grant College Program University of Alaska, Fairbanks (pp. 97-113).

Waples, R. S. (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, *89*(5), 438-450.

Waples, R. S., & Gaggiotti, O. (2006). INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular ecology*, *15*(6), 1419-1439.

Waples, R. S., Punt, A. E., & Cope, J. M. (2008). Integrating genetic data into management of marine resources: how can we do it better?. *Fish and Fisheries*, *9*(4), 423-449.

Ward, R. D., Woodwark, M., & Skibinski, D. O. F. (1994). A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of fish biology*, *44*(2), 213-232.

Webb, J. B., & Woodby, D. (2011). Long-term Alaska crab research priorities, 2011. *Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J11-04, Juneau*.

Weersing, K., & Toonen, R. J. (2009). Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series*, *393*, 1-12.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *evolution*, *38*(6), 1358-1370.

Weslawski, J. M., Pedersen, G., Petersen, S. F., & Porazinski, K. (2000). Entrapment of macroplankton in an Arctic fjord basin, Kongsfjorden, Svalbard. *Oceanologia*, *42*(1).

Westphal, M., & Nichols, B. (2018, October 2). Commercial Fisheries News Release, Bering Sea Tanner Crab Season Opens October 15 Total Allowable Catch Announced. *Alaska Department of Fish and Game, Division of Commercial Fisheries, Homer*.

Wood, K., Stratman, J. P., Palof, K., & Messmer, A. (2017). *Annual Management Report for the 2016/2017 Southeast Alaska/Yakutat Tanner Crab Fisheries*. Alaska Department of Fish and Game, Division of Sport Fish, Research and Technical Services.

Woodby, D., Carlile, D., Siddeek, S., Funk, F., Clark, J. H., & Hulbert, L. (2005). Commercial fisheries of Alaska. *Alaska Department of Fish and Game, Special Publication*, (05-09), 74.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*(1), 323-354.

Xuereb, A., Benestan, L., Normandeau, E., Daigle, R. M., Curtis, J. M., Bernatchez, L., & Fortin, M. J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RAD seq, in a highly dispersive marine invertebrate (Parastichopus californicus). *Molecular ecology*, *27*(10), 2347-2364.

Zheng, J., & Kruse, G. H. (2006). Recruitment variation of eastern Bering Sea crabs: climate-forcing or top-down effects?. *Progress in Oceanography*, *68*(2-4), 184-204.

Zheng, J. (2008). Temporal changes in size at maturity and their implications for fisheries management for eastern Bering Sea Tanner crab. *Journal of Northwest Atlantic fishery science*, *41*, 137-149.

Chapter 1 : Characterizing genetic diversity and population genetic structure of Tanner crab *Chionoecetes bairdi* in Alaska using reduced-representation sequencing

Abstract

Information on the extent of migration and connectivity of wild populations is an important component of the development of natural resource management policies. This study aims to develop baseline estimates of the genomic variation and population genetic structure of Tanner crab (*Chionoecetes bairdi*), a species that is important to Alaska's economy and culture. Tanner crab are the target of commercial, personal use, and subsistence fisheries in Alaska. Previously strong crab fisheries in Prince William Sound, Kodiak, and the Bering Sea have faced major uncertainties over the past few decades. Here, we investigate the population structure of *C. bairdi* in Alaska using a genotyping-by-sequencing approach. A set of 89 individuals, representing four sampling regions (Southeast Alaska, Prince William Sound, Bering Sea west of 166° W, Bering Sea east of 166° W) were sequenced using the double digest Restriction-Associated DNA sequencing (ddRADseq) method. We identified and genotyped 2,740 independent and neutral single-nucleotide polymorphisms (SNPs) and found evidence of high gene flow, with no support for subpopulation partitioning. The average observed heterozygosity and average expected heterozygosity estimates were similar across all sampled regions, with average observed heterozygosity (~16%) being significantly lower than average expected heterozygosity (~18%) in each sampling region. An analysis of molecular variance indicated that most of the observed variation occurred within individuals (90.53%), while 9.4 % was attributed to variation among individuals within regions, and no significant variation was measured among regions. All pairwise F_{ST} estimates between sampling regions were universally low, and we could not reject the null hypothesis of panmixia. Global F_{ST} was also low (0.000229), both in absolute magnitude and in comparison to estimates generated for other marine species.

Introduction

Delineating management regions to correspond with natural breaks in population demographic and/or genetic structure is important for effective management of fisheries and protection of marine organisms (Allendorf et al., 2010). Simulations demonstrate that over-splitting well-connected populations into separate stocks can lead to inefficient allocation of government resources to determine and enforce regulations and skew recruitment estimates, whereas failure to account for genetic structure in management schemes can lead to masking effects that severely impact or unknowingly eliminate distinct subpopulations (Spies & Punt, 2015; Spies, Spencer, & Punt, 2015). Accurately identifying regional population boundaries in marine organisms is particularly challenging because of the lack of obvious barriers to migration, especially for species with pelagic larval life stages (Palumbi, 2003; Sale et al., 2005). In many marine organisms, dispersal distance is only weakly correlated with pelagic larval duration, because physical characteristics such as water column stratification, shore topography, and tidal currents can affect larval advection and retention (Cowen & Sponaugle, 2009; Shanks, 2009; Weersing & Toonen, 2009). To circumvent the impracticalities of tagging microscopic larvae, estimates of genetic population structure have been useful for identifying regions with high connectivity from more isolated regions in species that have pelagic larvae (Barber et al., 2002; Benestan et al., 2015; Knutsen et al., 2011; Reitzel et al., 2013; Xuereb et al., 2018).

Genetic differentiation in marine organisms is usually much lower than in freshwater systems because of large effective population sizes spread across extensive geographic regions and high rates of migration between areas (Ward et al., 1994). Thus, identifying enough polymorphic loci to detect patterns of population structure in marine species presents a distinct challenge. Reduced-representation sequencing is gaining popularity in this field because it can be applied to non-model organisms and detect large numbers of genetic markers from across the genome, increasing the resolution of population genetic studies so that they are informative even when applied to organisms with very low levels of genetic differentiation (Peterson et al., 2012). These approaches have recently been used to resolve fine-scale structure and detect pathways of connectivity in other marine invertebrates (Rodríguez-Ezpeleta et al., 2016; Van Wyngaarden et al., 2017; Xuereb et al., 2018).

Tanner crab (*Chionoecetes bairdi*) is a benthic oregoniid crab that exhibits pelagic larval life stages lasting between two and three months (Haynes, 1973; Haynes, 1981; Incze et al., 1982; Urban & Hart, 1999). Upon reaching maturity, *C. bairdi* mate in the spring. They have been observed mating in large aggregatory mounds, and timing the release of larvae with tidal currents (Stevens et al., 1994; Stevens, 2003). Females possess sperm reservoirs and can retain sperm from previous annual mating seasons to fertilize clutches in subsequent years (Paul, 1984). Adults and juveniles tend to occupy non-overlapping areas, presumably to avoid conspecific predation (Nielsen et al., 2007). A focus on observations of any one of these life stages in isolation paints an incomplete picture of the degree of connectivity between *C. bairdi* across Alaska. Genetic data collected from a genome-wide scan of single nucleotide polymorphisms (SNPs) can provide an overall estimate of patterns of relatedness of *C. bairdi* throughout the Alaskan regions, and can be used to provide direction for specific observational studies or management considerations.

Chionoecetes bairdi is subject to a range of management concerns across the state. In Southeast Alaska, there is a high prevalence of infection by a dinoflagellate parasite, *Hemadotodinium*, which causes Bitter Crab Disease (BCD) or Bitter Crab Syndrome (Eaton et al., 1991). Information about the population structure or genetic diversity of aggregations within this region might help explain why some areas have a 95% prevalence of BCD while others appear to be unimpacted (Meyers et al., 1990). In the Eastern Bering Sea (EBS), overfishing limits and acceptable biological catch for *C. bairdi* are determined based on stock assessments of the total EBS region, but the total allowable catch is determined from two separate units east and west of 166° W. The biological relevance and validity of this distinction is a routinely questioned in annual stock assessment and fishery evaluation reports for Bering Sea Tanner crab fisheries (NPFMC, 2016; Stockhausen, 2018). The division between east and west of 166° W (represented in Figure 1.1) was originally based on differences in the size at maturity on either side of the boundary, and is supported by evidence of genetic differentiation based on allozyme data (Merkouris et al., 1998; Somerton, 1980; Somerton, 1981). Models of larval dispersal in the EBS built with the Regional Ocean Modeling System suggest that larvae produced by mature females in the eastern region of the EBS may be retained on that side of the boundary (Richar et al., 2015). However, there is not enough support to definitively confirm that these regions are

independently affected by harvests, so in some situations they are treated as a single unit, and in others they are considered independently. In 2015, the minimum preferred size-at-harvest in the eastern region was changed from 140 mm to 127 mm to match that of the west side. During the 2018/2019 season, the west side was open to commercial harvest and given a 2,439,000-lb harvest limit, while the east side remained closed for the season (Stockhausen, 2018). Abundance in the EBS and Prince William Sound (PWS) regions has fluctuated widely over time, and while this could be heavily influenced by environmental variation (Zheng & Kruse, 2006), these fluctuations may also reflect management boundaries that do not align with biologically relevant population structure to account for potential sources and sinks (Spies & Punt, 2015).

We applied reduced-representation sequencing methods to identify sites of genetic variation in *C. bairdi*, in an effort to elucidate details on the population genetic structure of this species within Alaskan waters. This study on the genetic population structure of *C. bairdi* contributes to the growing body of literature applying reduced-representation sequencing to non-model marine invertebrates to estimate levels of natural standing genetic variation and evaluate genetic population structure based on thousands of genome-wide SNP loci. It also provides insights that apply to specific management concerns in Alaska.

Methods

Sample collection

Adult *C. bairdi* crabs were sampled between the years of 2009-2016 from 4 regions: Southeast Alaska (SEAK), Prince William Sound (PWS), Bering Sea west of 166° W (EBS_NW), Bering Sea east of 166° W (EBS_SE). Crabs were collected by pots or trawl during annual stock assessment surveys by the National Marine Fisheries Service (NMFS) and the Alaska Department of Fish and Game (ADF&G) (Figure 1.1). Samples of tissue (dactyl segment of walking leg) or hemolymph (0.2 mL) were stored in 95% molecular grade ethanol at -20°C. A subset of 25 samples per region, were selected for DNA extraction and sequencing. Collection details are summarized in Appendix A.

DNA extraction, ddRADseq library preparation, and sequencing

Total genomic DNA was extracted using the reagents and protocols of the Gentra Puregene Tissue Kit from QIAGEN (QIAGEN, Valencia, CA, USA). Quality of isolated DNA was assessed by agarose gel electrophoresis. DNA quantity and purity was checked by fluorometry on a Qubit 2.0 instrument and by spectrophotometry on a Nanodrop 1000 instrument, respectively. Samples were processed according to Double Digest Restriction Associated DNA sequencing (ddRADseq) protocol as described in Peterson et al., 2012. ddRADseq is a reduced-representation sequencing strategy to target sequencing effort to a subset of genomic fragments bounded by restriction enzyme target motifs. Genomic DNA from each sample was fragmented using a combination of the restriction enzymes EcoRI-HF (target motif: GAATTC) and MspI (target motif: CCGG) to generate pools of variable length fragments flanked by the restriction enzyme recognition sites. The fragment pools were appended with Illumina sequencing adapters and identification indices so that samples could be multiplexed for sequencing. The tagged genomic fragment pools, termed libraries, were size-selected at 350-400bp, using a BluePippin (Sage Science, Beverly, MA, USA). The libraries were pooled and sequenced on two lanes of an Illumina Hi-Seq platform (Illumina Inc, San Diego, CA, USA) to generate > 1 million paired-end 150 base pair (bp) sequence reads from each individual. Library preparation, multiplexing, and sequencing were performed by Research and Testing Laboratory (RTL; Research and Testing Laboratory, Lubbock, TX, USA). Reads were demultiplexed and tags were removed in-house at RTL.

Read quality control, SNP discovery, and genotyping

Demultiplexed sequencing files were examined using the quality assessment modules implemented in *FASTQC* (Andrews, 2010) to assess overall sequence quality and to obtain summary statistics. The demultiplexed sequences were filtered, clustered, and aligned *de novo* using the *ipyrad* pipeline (v. 0.7.28; Eaton & Overcast, 2016) to call variant sites and assemble a dataset of individual genotypes. A range of inputs for the parameters of minimum read depth and clustering thresholds were tested to check for potential bias in SNP calls. Adapters and restriction cut-site overhangs were identified and trimmed from reads before filtering. Reads with more than 5 low-quality bases (< 33 phred Q score offset), and/or a read length less than 35 bp were filtered out before clustering and alignment. Reads were clustered *de novo* within

individuals at a threshold of 85% similarity. Clusters with a sequencing depth < 10 or > 40,000 reads were discarded. Then, clusters were scanned for adherence to a diploid model, and consensus sequences were evaluated. Consensus sequences with more than 5 uncalled bases were filtered out. These consensus sequences were clustered across individuals at a threshold of 85% similarity to identify putative loci. Loci that contained > 20 SNPs within either read 1(R1) or read 2 (R2), contained > 8 insertions or deletions within R1 or R2, contained heterozygous sites that were shared by > 50% of individuals, or were sequenced in less than 4 individuals were discarded. Variant sites within the locus set were identified, and individual genotype calls at those sites were organized into a genotype assembly and output to a variant call format (VCF) file. Then *vcftools* (v.0.1.5, Danecek et al., 2011) was used to remove insertions and deletions, filter out SNP sites with a minor allele frequency (MAF) < 0.03 or minor allele count (MAC) < 3, and remove sites that were genotyped in fewer than 50% of individuals. The genotype assembly was further filtered using a thinning threshold of 300 bp to retain only the first SNP from each locus (150bp from R1, 150 bp from R2) to limit linkage and non-independence in the dataset. The genotype assembly was imported to *R* (R Core Team, 2017) using the *vcfR* package (Knaus & Grünwald, 2017) and SNPs with more than two alleles were removed. To minimize the proportion of missing data, loci that were genotyped in fewer than 80% of individuals and individuals that were genotyped at fewer than 80% of remaining loci were removed using tools from the *R* package *dartR* (Gruber et al., 2018). Each locus was tested for adherence to Hardy Weinberg expectations (HWE) within each of the four sampling regions using a Monte Carlo test and a Bonferroni-corrected test statistic using the *R* package *dartR* (Gruber et al., 2018). Loci that deviated from HWE in any region were discarded from the genotype assembly. Locus neutrality was assessed using the *outflank* implementation in the *R* package *dartR*, with a q-value false discovery rate threshold of 0.05 (Gruber et al., 2018; Whitlock & Lotterhos, 2015). Loci potentially under selection were identified and removed from the assembly.

Genetic diversity and population genetic analyses

The *R* package *adegenet* (v.2.1.1, Jombart & Collins, 2015) was used to calculate and plot summary information for the genotype assembly. Heterozygosity estimates at each SNP site were averaged to calculate mean observed heterozygosity and expected heterozygosity under HWE for each sampling region. A paired t-test was used to evaluate whether observed

heterozygosity significantly differed from expected heterozygosity in each sampling region. An analysis of molecular variance (AMOVA) implemented in the *R* package *poppr* (Kamvar et al., 2014) was used to estimate variations within individuals, subpopulations, and within the total population, and to estimate Φ statistics. The estimates were tested for significance using a Monte Carlo test with 1,000 permutations with the *R* package *ade4* (Chessel et al., 2009; Dray & Dufour, 2007; Excoffier et al., 1992). Global F_{ST} was also estimated according the Weir & Cockerham (1984) method implemented in the *R* package *hierfstat* (Goudet, 2005). Pairwise Nei's F_{ST} estimates were calculated using the *R* package *adegenet* (v.2.1.1, Jombart & Collins, 2015; Nei, 1973) and significance was assessed with a Monte Carlo test with 1,000 permutations, at $\alpha = 0.05$ while correcting for multiple comparisons $k * \alpha/t$, where k is the ordered p value rank, α is the significance level, and t is the number of tests, using the *R* package *ade4* (Chessel et al., 2009; Dray & Dufour, 2007).

A principal components analysis (PCA) on SNP genotypes was performed using *adegenet* (v.2.1.1, Jombart & Collins, 2015) to identify potential clustering patterns among individuals. To test for possible subpopulation clustering, a model-free test of sequential K-means was done through a discriminant analysis of principal components (DAPC) using the *R* package *adegenet* (v.2.1.1, Jombart & Collins, 2015), and Bayesian Information Criterion (BIC) was used to assess the best supported model of clustering (Jombart et al., 2010). Then a DAPC was used with our predefined sampling regions to test individual assignment accuracy to recognized management regions using tools from the *R* package *adegenet* (v.2.1.1, Jombart & Collins, 2015). The appropriate number of principal components retained for the discriminant analysis was determined by cross-validation with 1000 permutations. Population structure was tested using the Bayesian clustering approach implemented in the program *STRUCTURE* (v.2.3.4, Pritchard et al., 2000). The program *StrAuto* (v.1.0, Chhatre & Emerson, 2017) was used to automate 10 replicates for each K tested, 1 - 4, with 100,000 iterations after 10,000 burn-in iterations. The results of the replicate tests were summarized using the CLUMPAK server (Kopelman et al., 2015).

Results

Sequencing

Of the 100 DNA extraction samples that were processed and sequenced by RTL Genomics, 93 individual samples yielded > 0.5M reads. The 7 individuals that had < 0.5M reads were dropped from the analysis. FastQC reports on the overall sequence quality, read length, and adapter content indicated that the sequencing run met criteria necessary to proceed with downstream analyses. Within the final dataset, the number of reads per individual ranged from 605,824 to 4,793,785, with an average of 1,897,727 reads per individual after quality filtering. On average, only 0.876% of raw reads were filtered out due to low quality. Results from filtering and assembly steps are reported in Table 1.1.

SNP discovery and genotyping

An average of 93,673 sequence clusters were identified within each individual. After cluster filtering, 81,346 raw loci were identified (a locus referring to a ~150bp sequence region). From the 21,211 loci retained after filtering, *ipyrad* identified 135,636 putative variant sites (including SNPs, insertions, and deletions). Most of the variant sites were only observed in one individual, and were likely erroneous. After filtering by MAF, MAC, coverage, neutrality thresholds, and the first SNP from each locus, 3,707 SNPs were retained. Four individuals were dropped because they were sequenced at less than 80% of loci. Final sample sizes for each of the sampling regions are reported in Table 1.2). All of the SNPs met Hardy-Weinberg expectations in each of the four regions samples at $\alpha = 0.05$ with Bonferroni correction. Three SNP sites were identified as potentially under directional selection and removed, leaving 2,740 SNPs used for F_{ST} hypothesis testing (Table 1.1).

Genetic diversity estimates

Heterozygosity estimates by sampling region are reported in Table 1.2. Observed heterozygosity within sampling regions ranged from 16.1% to 16.6% while expected heterozygosity based on HWE ranged from 17.8% - 18.4%. In each sampling region, observed heterozygosity was significantly lower than the expected heterozygosity under HWE. This can also be observed in Figure 1.2, a scatterplot of observed versus expected heterozygosity for all SNP sites. The AMOVA indicated that most of the variance observed was at the individual level (Table 1.3).

Only the variation within individuals and the variation between individuals within regions was significant, and accounted for approximately 90.5 % and 9.42 % of total variation, respectively.

The estimated excess similarity of alleles within an individual relative to the population (Φ_{IT}) was 0.0947. The estimated excess similarity of alleles within an individual relative to its sampling region (Φ_{IS}) was 0.0943. The estimated excess similarity of alleles within a sampling region relative to the total population or proportion of genetic diversity due to allele frequency differences among sampling regions (Φ_{ST}) was 0.000456. The analogous statistic, global F_{ST}, was estimated to be 0.000229.

Population differentiation estimates and population structure

Pairwise F-statistic estimates indicated that there is no significant genetic differentiation among crabs from the sampling regions included in this study. Observed pairwise F_{ST} values did indicate more similarity between individuals from the two Bering Sea sampling regions, and progressively more differentiation when compared to PWS and Southeast Alaska (Figure 1.3), but these observations were not statistically significant based on Monte Carlo tests (Figure 1.4). Observed pairwise F_{ST} values ranged from 0.0122 to 0.0145, and we could not reject the null model of panmixia for any of the pairwise estimates based on Monte Carlo tests. Results from F_{ST} analyses are summarized in Table 1.4.

The results of PCA were consistent with the results obtained by F_{ST} calculations and Monte Carlo tests, in that they did not indicate evidence of subpopulations. There was substantial overlap between all of the sampling regions when plotting the first two principal components (Figure 1.5) from PCA with 3 axes retained. DAPC clustering tests of K-means without population priors did not support any subpopulation clustering, as the lowest BIC was reported for K=1 cluster. DAPC assignment tests for predefined regions with a training set of 60 individuals and testing set of 29 individuals were, on average, only able to accurately assign 28.1% of the testing set individuals to their true collection region (Figure 1.6, Figure 1.7), and did not differ from the 95% confidence interval of assignment success by random chance (95% Confidence Interval for random chance was (17.6, 34.9)). Bayesian clustering analysis also indicated the same lack of detectable genetic population structure. We tested up to four

population partitions with 10 replicate tests each, and did not find support for K = 2, K = 3, or K = 4 subpopulations. Mean posterior probabilities favored K = 1 (Mean LnP(K) = -152260, SD = 82.05), and decreased for higher values of K. Delta K was highest for K = 3, but delta K cannot test support for K = 1, and the summary plots of individual Q values clearly favored K = 1. The summary plots are representative of high gene-flow scenarios, where individuals are unlikely to be assigned to any particular cluster, and instead have equivalent assignments to each (Figure 1.8).

Discussion

This is the first study to identify variation at thousands of sites within the Tanner crab genome and to apply those data to quantify genetic population structure. Analysis of genome-wide SNPs did not indicate significant genetic population structure in Tanner crab in Alaskan waters. We did not detect significant population differentiation through Monte Carlo tests of Nei's pairwise F_{ST} among established management regions. Bayesian clustering analysis with STRUCTURE also did not support any partitioning of sampled individuals into separate groups, and PCA demonstrated a lack of genetic differentiation among Tanner crab samples by region. This supports the hypothesis that the long larval period of this species is conducive to substantial gene flow between regions.

These results conflict with previous analyses of Tanner crab genetic population structure. Variation in allozymes indicated statistically significant differentiation between the Bering Sea, Gulf of Alaska, and Southeast Alaska, as well as significant differences between samples collected east and west of 166°W in the Bering Sea (Merkouris et al., 1998). Notably, those estimates were based on fewer markers, and the differentiation could very well be due to differential selection pressures between regions rather than restricted gene flow. In other studies where DNA differentiation estimates were much lower than differentiation based on allozymes, results were interpreted to mean that the allozymes used were not selectively neutral (Lemaire et al., 2000), or that their observed variation was driven by phenotypic plasticity rather than mutation, drift, and/or selection (Olsen et al., 2014). Differences in size at maturation are a documented phenotypic difference between *C. bairdi* east and west of 166° W (Somerton, 1980), but factoring in our results, this does not look like a case of genetic drift occurring in isolated

populations. The degree to which these phenotypic differences could be due to phenotypic plasticity or selection has not been evaluated. Either of these phenomena, if associated with the proteins used in allozyme assays, could lead to different conclusions than found with this set, composed of neutral SNPs (Lemaire et al., 2000; Olsen et al., 2014). The closely related species, *C. opilio,* also varies in size along latitudinal cline that is associated with a temperature gradient, (Burmeister & Sainte-Marie, 2010), yet panmixia of *C. opilio* throughout its Alaskan range is also supported by genetic analysis (Albrecht et al., 2014).

Sometimes population differentiation estimates in marine species are low because of high gene flow, large effective population sizes, or recent population origin (Hellberg et al., 2002; Ward et al., 1994). Another factor leading to low subpopulation-level structure could be high levels of genetic variation within individuals, which leaves less gene diversity that can be attributed to differences in allele frequencies among subpopulations (Hedrick, 1999). This was the case in the present study. We observed most of the variation at the level of individuals, and the remaining variation was attributed to differences among individuals within regions, while no significant variation was attributed to differences between regions. Indeed, the homogeneity among regions led to unsuccessful assignment tests, because there were not enough variations unique to each sampling region to effectively assign unidentified individuals back to their respective regions. Accurate assignment is usually possible in species like salmon (Hess et al., 2011), which have highly differentiated stocks, but is less feasible in marine organisms that exhibit these low levels of regional differentiation, and is not possible in this study with the collected set of unlinked, neutral SNPs. Assignment capabilities are sometimes improved by using non-neutral markers (Ackerman et al., 2011). The selective pressures acting in different regions can have detectable differences in allele frequencies, which have led to successful assignment even in marine invertebrates with pelagic larval periods (Benestan et al., 2015). We only detected three putative SNPs under directional selection, so the current dataset does not contain enough non-neutral loci to test this capability.

The failure to reject the null hypothesis of panmixia is not due to a lack of statistical power. According to simulations by Patterson et al. 2006, we can calculate the threshold of differentiation that we have the sampling power to detect. $Threshold\ F_{ST} = {}^{1}/_{\sqrt{mn}}$ where m is

the number of markers (2740 in the present study) and n is the number of individuals sampled from subpopulations (lowest n = 21 for present study). Thus, we should have the statistical power to detect an F_{ST} of 0.004 between sets of regions. Yet, our estimates of regional differentiation, F_{ST}, were at the level of 0.01, and we still did not detect significant differences or find support for subpopulation clustering through eigenanalysis with DAPC. Thus, our results should not be interpreted as a lack a sufficient number of markers or individuals, but as actual low differentiation among regions. Nevertheless, these results could be impacted by sampling design (Puechmaille, 2016). The regions included in this study are not representative of the full range of the species in Alaskan waters, so this dataset may not include all potential source populations. The Global F_{ST} estimate thus does not reflect all possible subunits.

In addition to testing for population structure, we also comprehensively examined the genetic diversity of Tanner crab in Alaska. Genetic diversity is used as an indicator of a population's fitness and resilience to disturbances (Hughes et al., 2008). Heterozygosity estimates and observations fell within ranges observed in other marine invertebrate RADseq analyses (Reitzel et al., 2013; Xuereb et al., 2018). On average, the observed heterozygosity at each locus (~ 0.16) was 2% lower than the heterozygosity we would expect under Hardy-Weinberg conditions (~ 0.18), and this difference was statistically significant in each of our sampling regions. Negative heterozygosity ratios (i.e., heterozygote deficiencies) are often used as a way to identify and evaluate inbreeding, but this observation could also result from sampling effects or genotyping artifacts (Lynch, 2008; Sinnock, 1975).

Two ways that the sampling scheme could impact our heterozygosity estimates are through Wahlund effects (Sinnock, 1975) or by incomplete representation within sampling regions (Robertson, 1965). If a sampling region actually contains two or more subpopulations that possess different allele frequencies, measured overall heterozygosity is reduced. The Wahlund effect is probably not responsible for our observations because if two or more subpopulations were truly present in our regional samples, our clustering analyses should have detected subpopulation partitioning. Our sample sizes within regions were relatively small. If these small sample sizes were inadvertently collected from areas where individuals are more closely related, they may not accurately represent the scope of diversity in the region and lead to lower estimates

of heterozygosity (Robertson, 1965). This is not likely a concern for our samples because the overall average heterozygosity we measured matched with the estimate within each sampling region, and there was very little variance among those estimates ($\sigma = 0.002$). If sampling was not representative, we would expect higher variance in the estimates among different regions.

Genotyping artifacts may also impact estimates of heterozygosity either due to insufficient sequencing depth or through overrepresentation of cloned sequences from polymerase chain reactions (i.e., PCR duplicates) represented in the sequence files (Lynch, 2008; Puritz et al., 2014). Tests of the impact of sequencing depth and missing data on heterozygosity estimates for another marine invertebrate found that estimates of heterozygosity were consistent when the sequencing depth was 10 reads or greater, and missing data had no impact on the estimates (Lal et al., 2016). We used a threshold sequencing depth of 10 reads for each of the genotypes in our final assembly, so the excess in homozygosity is probably not due to insufficient sequencing depth. Our estimates of heterozygosity could be inflated by PCR duplicates that were not filtered out during our assembly steps. One downfall of the ddRADseq method is that because the fragments have identical starting positions and lengths, it is not possible to identify and remove PCR duplicates. Theoretically, retaining sequence reads that are clonally created during the PCR stages can give a false read depth and upwardly bias estimates of homozygosity, but this has not been empirically tested (Puritz et al., 2014). We cannot rule out the possible bias due to PCR duplicates.

If the heterozygote deficiency is not caused by sampling or genotyping biases, the results can be interpreted as measures of inbreeding or alleles that are identical-by-descent (Holsinger & Weir, 2009; Wright, 1950). Heterozygosity estimates from before and after a period of commercial fishing overexploitation of *Pleuronectes platessa* in the North Sea demonstrated heterozygote deficiency following the period of overexploitation. The estimated effective population size in the region was still large enough that inbreeding should not have had a substantial effect on genetic diversity, so inbreeding was hypothesized to be a behavioral preference triggered by fishing disruption of spawning habitat (Hoarau et al., 2005). We cannot make this assertion, as we do not have heterozygosity estimates from before fishery exploitation of Tanner crab stocks or knowledge of behavioral responses to fishing. A reasonable explanation for our detection of

inbreeding could be the hypothesis of Sweepstakes Reproductive Success (SRS), which is postulated to affect species with low parental care and pelagic larval life stages. SRS occurs when stochastic fluctuations in the environment allow only the offspring of a small proportion of mature individuals to recruit. Thus, effective population size is much lower than the census size in these scenarios, which produces inbreeding signals in the genetic data because only a small proportion of the individuals are contributing to each generation (Hedgecock & Pudovkin, 2011). *Chionoecetes bairdi* meet the life-history characteristics associated with SRS, so this could be a factor contributing to the genetic diversity we observe.

Finally, the sequence data collected through this project may lend itself to further investigations of the species demography. This study relied on a dataset of independent, neutral SNP genotypes, but linked genetic variabilities in the form of microhaplotypes may encode more traceable information on population histories (Baetscher et al., 2018; Kidd et al., 2014; Lawson et al., 2012). Tools for utilizing microhaplotype information from the sequences generated through ddRADseq studies such as this one are emerging (Hendricks et al., 2018). Mitochondrial DNA haplotypes, in the past, have provided weak, inconclusive evidence supporting the hypothesis of *C. bairdi* larval advection by the Alaska Coastal Current (Bunch et al., 1998). Our data could support a repeated study with hundreds to thousands of haplotypes, rather than tens, as the ever-growing bioinformatics sector builds the resources to answer this and other questions.

Table 1.1: Individuals, loci, and variable sites retained following each filtering step

Step	Individuals Retained	Loci Retained	Variant Sites Retained
Samples sequenced at RTL	100	-	-
Remove extremely low coverage samples	93	-	-
Ipyrad			
Total prefiltered loci	93	81,346	-
Remove duplicates	93	79,193	-
Remove Insertions/deletions >8	93	73,653	-
Remove # SNPS >20	93	72,239	-
Remove shared heterozygosity >50%	93	69,104	-
Remove sample coverage <4 individuals	93	29,946	-
Raw variant site output from ipyrad	93	27,211	135,636
vcftools			
Remove insertions/deletions, MAF > 0.03, MAC >3, thin = 300 (to retain 1 SNP per locus)	93	-	3,816
R			
Remove non-binary SNPs	93	-	3,707
Remove sites genotyped in <80 % of individuals	93	-	2,743
Remove individuals with < 80% coverage	89	-	2,743
Remove Bayescan outliers	89	-	2,740
Remove sites out of HWE w/ Bonferroni correction	**89**	-	**2,740**

Table 1.2: Heterozygosity estimates

Heterozygosity estimates computed using the summary function of the R package adegenet and averaged within each sampling region. The sample size for each region, mean expected heterozygosity for loci within each sampling region according to HWE, the mean observed heterozygosity within each sampling region, the difference between observed and expected heterozygosity, and the p value from a paired t test testing if the observed value was different from the expected. All regions had significantly lower observed heterozygosity than expected under HWE.

Region	n	H_{exp}	H_{obs}	$H_{obs} - H_{exp}$	p value
EBS_NW	23	0.1824368	0.1627065	-0.0197303	< 2.2e-16
EBS_SE	24	0.183295	0.1652526	-0.0180423	< 2.2e-16
PWS	21	0.1781354	0.160684	-0.0174514	< 2.2e-16
SEAK	21	0.1844682	0.165884	-0.0185842	< 2.2e-16

Table 1.3: Results from an AMOVA

AMOVA Monte Carlo test with 1000 replicates. Only the variation within individuals and the variation between individuals within regions was significant, and accounted for approximately 90.5 % and 9.42 % of total variation, respectively.

	Sigma	%	Alter	P value
Variations between subpopulations	0.1815736	0.04544716	greater	0.18381618
Variations between samples within subpopulations	37.6577493	9.42558833	greater	0.000999
Variations within samples	361.687455	90.5289645	less	0.000999
Total variations		100		

Table 1.4: Results from Monte Carlo tests of pairwise F_{ST} estimates

Observed pairwise F_{ST} is reported above the diagonal. Below the diagonal is the p-value based on 1000 permutations to generate the null distribution. All of the pairwise estimates of F_{ST} between sampling regions were low and none were significantly different from 0, the assumption under panmixia, at $\alpha = 0.05$ after correction for false discovery rate with multiple comparisons.

	EBS_NW	EBS_SE	PWS	SEAK
EBS_NW	-	0.01219071	0.0135067	0.01453057
EBS_SE	0.96503497	-	0.01297118	0.01301969
PWS	0.35764236	0.68231768	-	0.01427561
SEAK	0.04595405	0.67332667	0.1028971	-

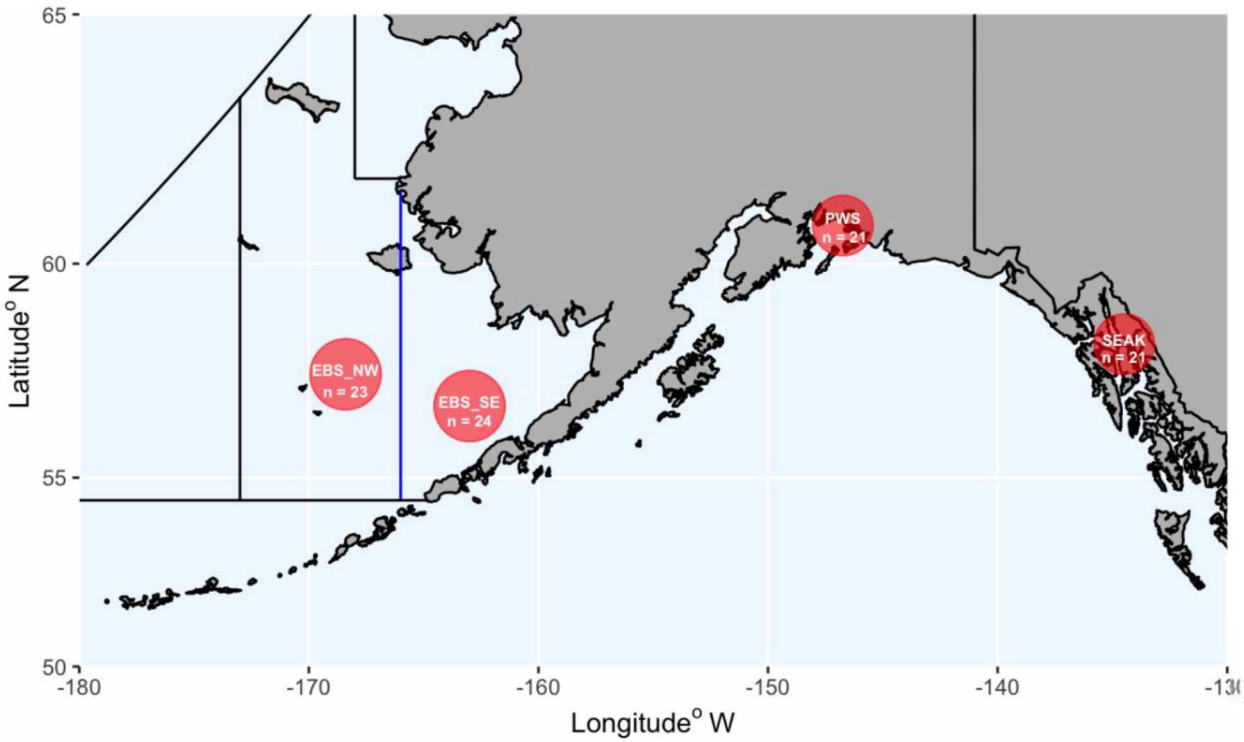

Figure 1.1: Map of sample collection regions.

Map of Alaska with the Bering Sea Tanner crab fishery borders marked by black lines, with 166°W marked in blue. Red circles indicate sampling regions. Each is labeled by the region code, and the sample size listed is the number of individuals from that region that are included in the final genotype assembly. Both sampling regions in the Eastern Bering Sea are estimates of where samples were collected, because those collections do not have geolocation data.

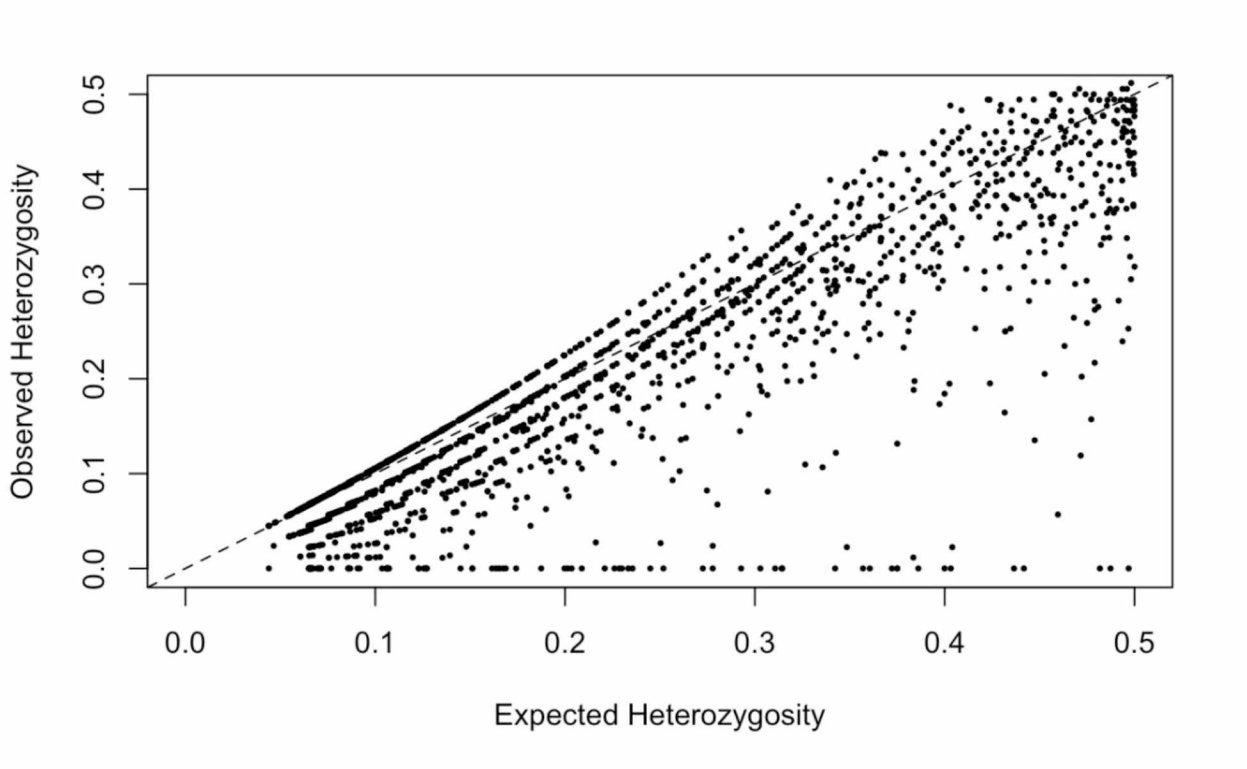

Figure 1.2: Scatter plot of observed vs expected heterozygosity for each SNP site

Scatter plot of the observed versus expected heterozygosity at each SNP site. Observed heterozygosity at most SNP sites was lower than the expected heterozygosity under HWE.

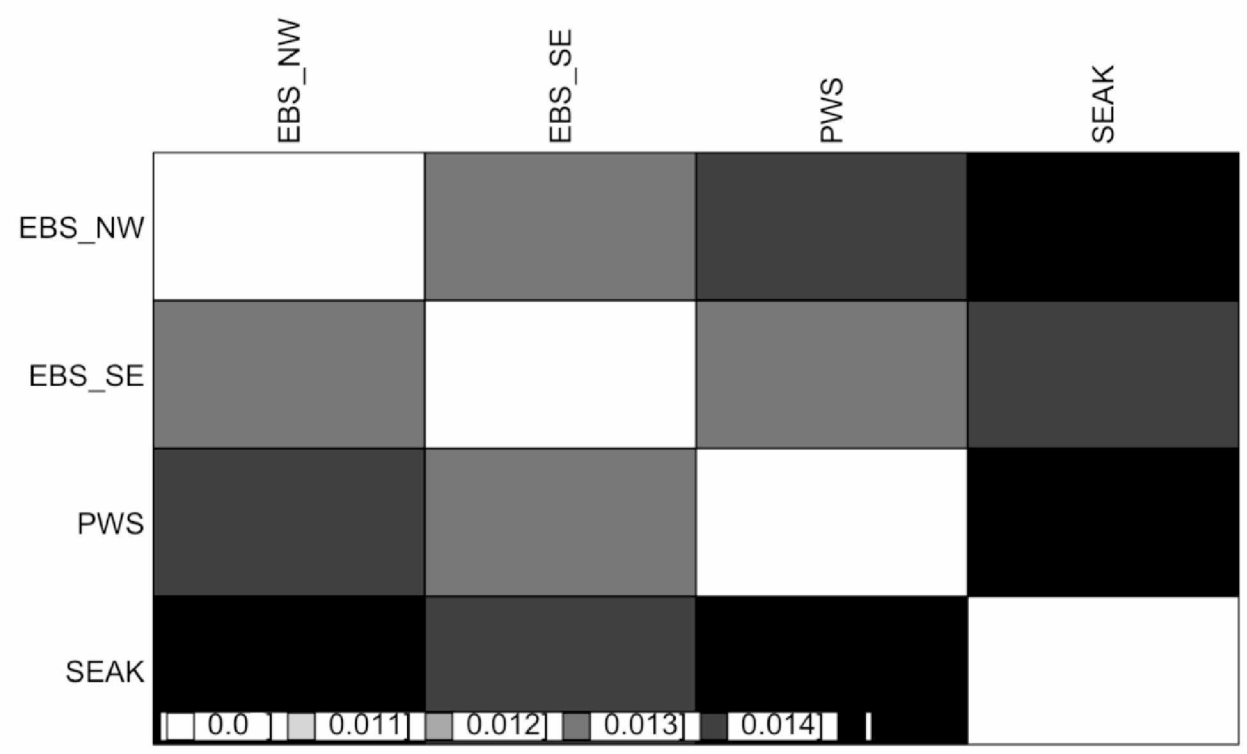

Figure 1.3: Heatmap of observed pairwise F_{ST} values

Heatmap of observed pairwise F_{ST} for each pair of sampling regions. The sampling regions are ordered from West to East. Darker values indicate larger observed F_{ST} and more genetic differentiation.

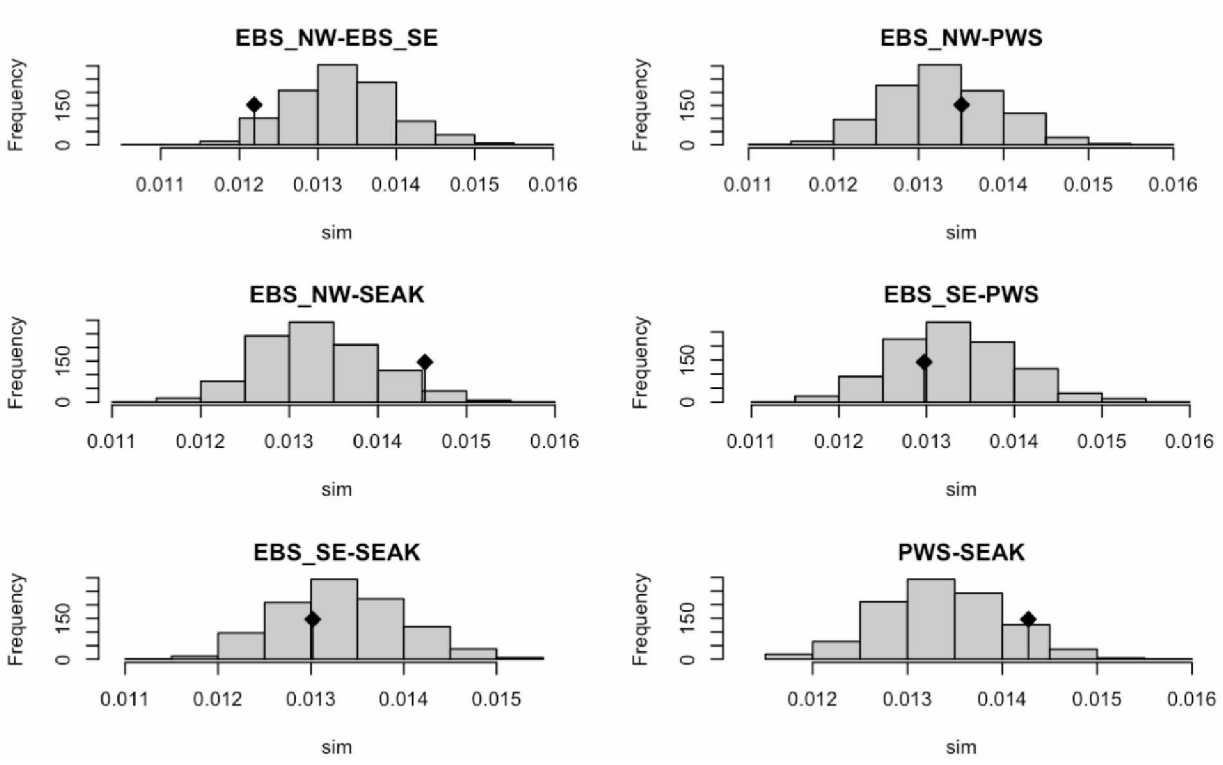

Figure 1.4: Permutation distributions from Monte Carlo tests

Plots of observed pairwise F_{ST} and the null distribution based on Monte Carlo tests for each pair of regions. Tests were performed with 1000 permutations of 89 individuals. The observed F_{ST} for each pair is marked by a black diamond. No pairwise F_{ST} estimates were significantly different than the null model of panmixia after correction for false discovery rate.

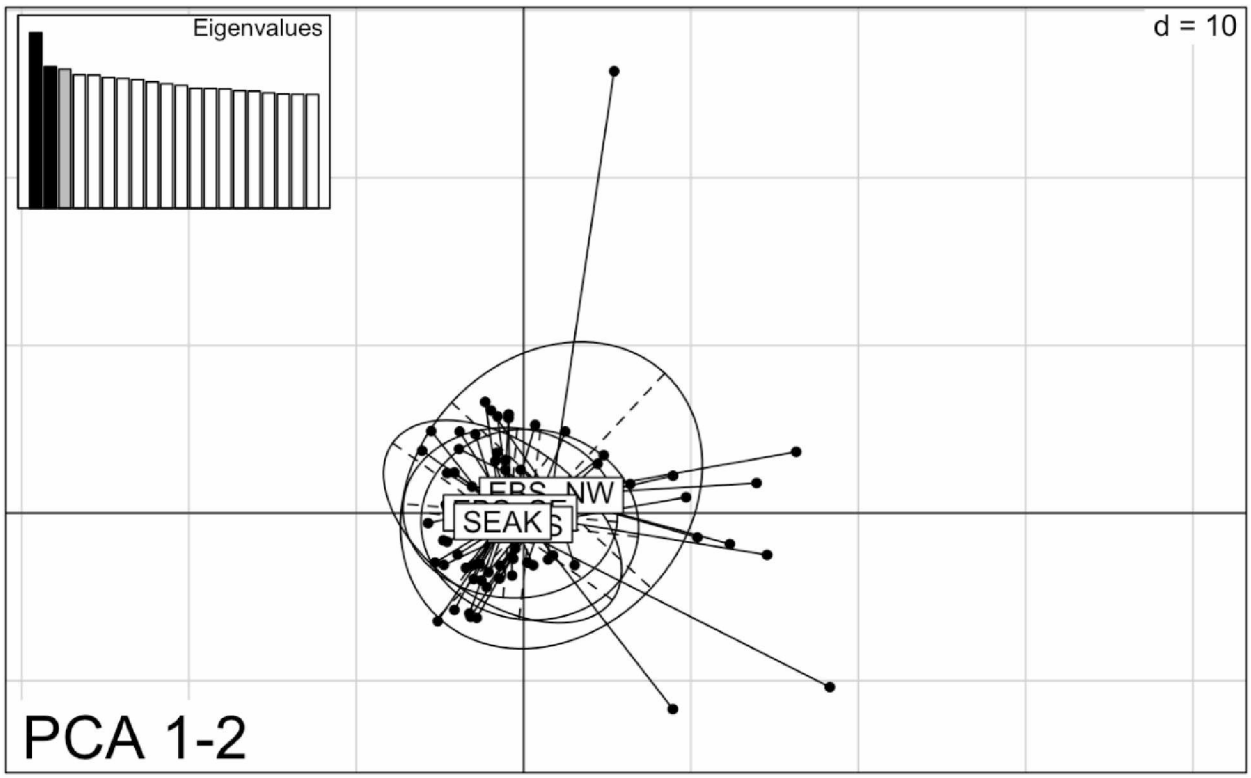

Figure 1.5: Plot of principal components 1 and 2 from PCA

PCA retaining 3 axes, with variation from the first principal component plotted along the x axis and variation from the second principal component plotted along the y axis . Eigenvalues 25.29453 (2.1 %), 20.40524 (1.7 %).

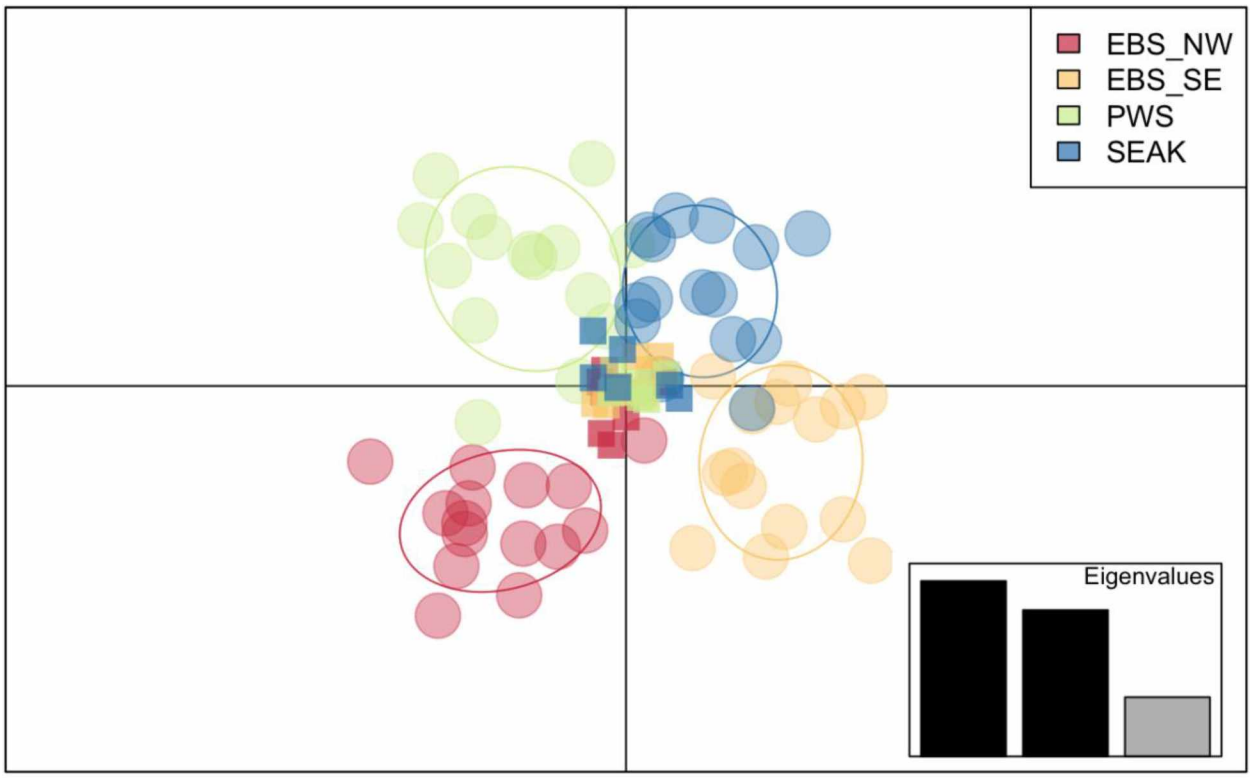

Figure 1.6: Plot of DAPC

Plot of iteration of DAPC with 41 retained axes in the PCA step and 3 retained axes in the Discriminant analysis step. Before DAPC, the dataset was split into a training set and testing set. The 60 individuals in the training set are represented by circles, and the assignment test placement of the 29 individuals used as the testing set represented by squares. All icons are shaded according to their true sampling region.

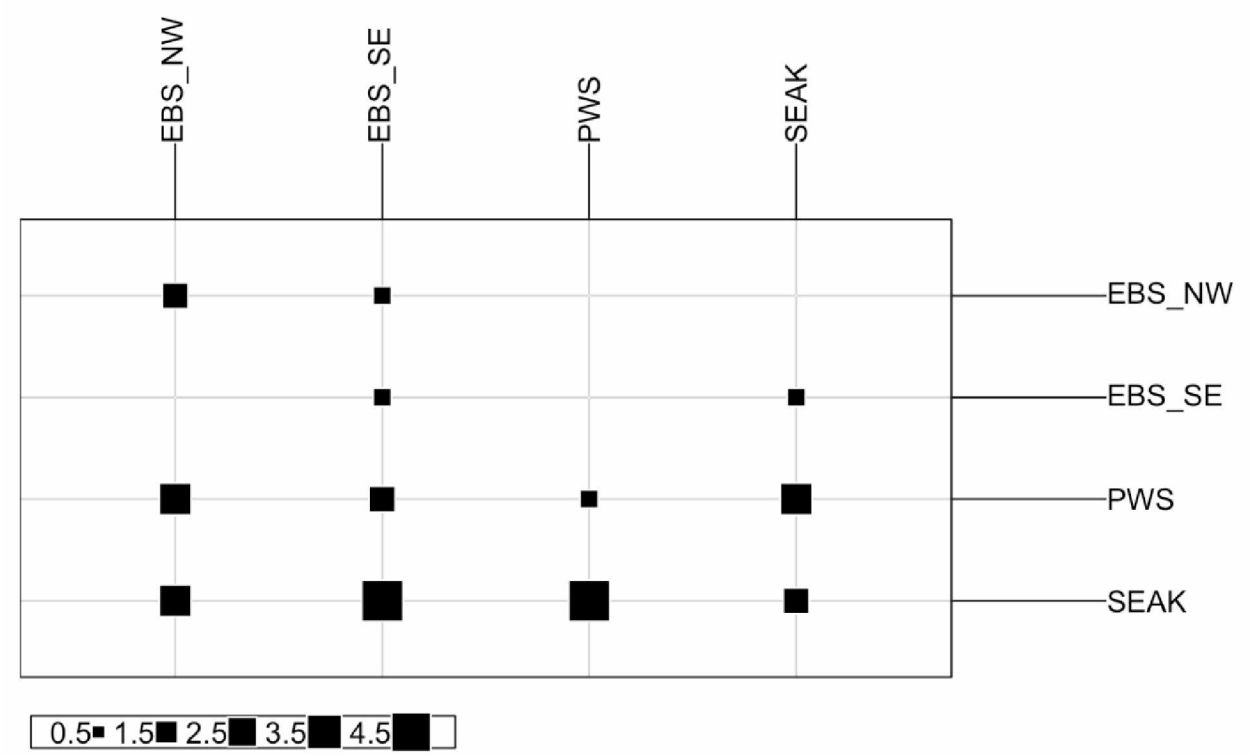

Figure 1.7: Assignments of 29 individuals by DAPC

Results from the assignment of 29 individuals that were withheld from one iteration of DAPC and used to test assignment accuracy. Only 20.6 % of the individuals were assigned to the correct sampling region in this iteration. Over 1000 iterations, mean assignment accuracy was 28.1%, and the 95 % confidence interval for random chance was (17.6 %, 34.9 %).

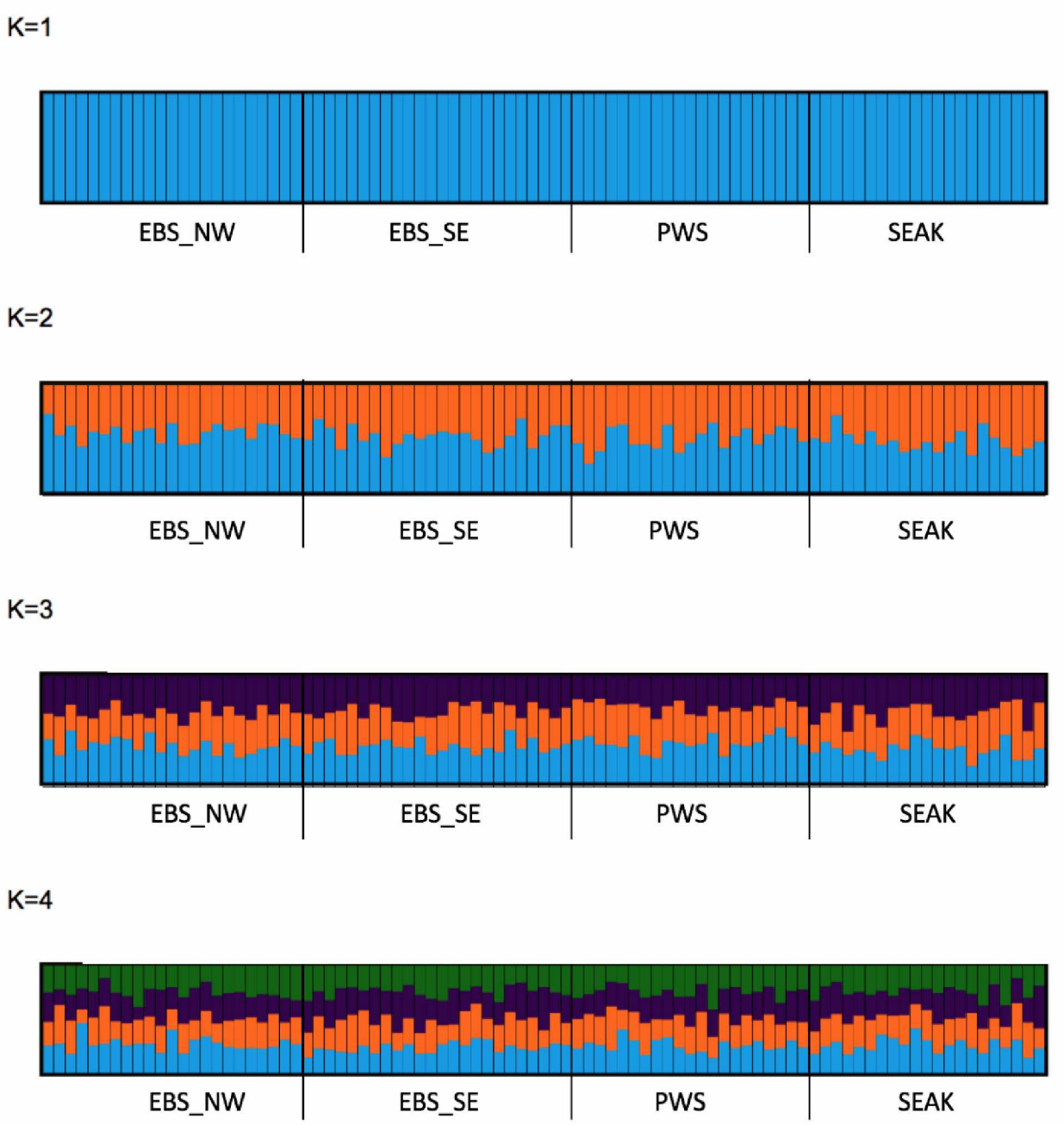

Figure 1.8: CLUMPAK plots of STRUCTURE analysis

CLUMPAK plots from STRUCTURE analysis testing K = 1 - 4 with 10 replicates each. There was no support for 2, 3, or 4 subpopulations. Each bar represents a sample, and they are ordered by geographic regions west to east.

References

Ackerman, M. W., Habicht, C., & Seeb, L. W. (2011). Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River, Alaska. *Transactions of the American Fisheries Society*, *140*(3), 865-881.

Albrecht, G. T., Valentin, A. E., Hundertmark, K. J., & Hardy, S. M. (2014). Panmixia in Alaskan populations of the snow crab Chionoecetes opilio (Malacostraca: Decapoda) in the Bering, Chukchi, and Beaufort Seas. *Journal of Crustacean Biology*, *34*(1), 31-39.

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature reviews genetics*, *11*(10), 697.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Cambridge, UK: Babraham Institute.

Baetscher, D. S., Clemento, A. J., Ng, T. C., Anderson, E. C., & Garza, J. C. (2018). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources*, *18*(2), 296-305.

Barber, P. H., Palumbi, S. R., Erdmann, M. V., & Moosa, M. K. (2002). Sharp genetic breaks among populations of Haptosquilla pulchella (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology*, *11*(4), 659-674.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). *Molecular ecology*, *24*(13), 3299-3315.

Bunch, T., Highsmith, R. C., & Shields, G. F. (1998). Genetic evidence for dispersal of larvae of Tanner crabs (Chionoecetes bairdi) by the Alaskan Coastal Current. *Molecular Marine Biology and Biotechnology*, 7(2), 153-159.

Burmeister, A., & Sainte-Marie, B. (2010). Pattern and causes of a temperature-dependent gradient of size at terminal moult in snow crab (Chionoecetes opilio) along West Greenland. *Polar Biology*, *33*(6), 775-788.

Chessel, D., Dufour, A. B., & Dray, S. (2009). Analysis of Ecological Data: Exploratory and Euclidean methods in Environmental sciences. *Version 1.4-14. 2 October 2010; 1: 4, 11.*

Chhatre, V. E., & Emerson, K. J. (2017). StrAuto: automation and parallelization of STRUCTURE analysis. *BMC bioinformatics*, *18*(1), 192.

Cowen, R. K., & Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Marine Science*, *1*.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.

Dray, S., & Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software*, *22*(4), 1-20.

Eaton, W. D., Love, D. C., Botelho, C., Meyers, T. R., Imamura, K., & Koeneman, T. (1991). Preliminary results on the seasonality and life cycle of the parasitic dinoflagellate causing bitter crab disease in Alaskan Tanner crabs (Chionoecetes bairdi). *Journal of Invertebrate Pathology*, *57*(3), 426-434.

Eaton, D. A. R., & Overcast, I. (2016). ipyrad: interactive assembly and analysis of RADseq data sets.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479-491.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184-186.

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, *18*(3), 691-699.

Haynes, E. (1973). Description of the prezoea and stage I zoea of Chionoecetes bairdi, and C. opilio (Oxyrhyncha, Oregoniinae). *Fishery Bulletin*, *71*(3), 769-775.

Haynes, E. (1981). Description of the stage II zoea of snow crab, Chionoecetes bairdi, (Oxyrhrncha, Majidae), from plankton of lower Cook Inlet, Alaska. *Fishery Bulletin*, *79*(1), 177-182.

Hedgecock, D., & Pudovkin, A. I. (2011). Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science*, *87*(4), 971-1002.

Hedrick, P. W. (1999). Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, *53*(2), 313-318.

Hellberg, M. E., Burton, R. S., Neigel, J. E., & Palumbi, S. R. (2002). Genetic assessment of connectivity among marine populations. *Bulletin of marine science*, *70*(Supplement 1), 273-290.

Hendricks, S., Anderson, E. C., Antao, T., Bernatchez, L., Forester, B. R., Garner, B., Hand, B. K., Hohenlohe, P. A., Kardos, M., Koop, B., Sethuraman, A., Waples, R. S., & Luikart, G. (2018). Recent advances in conservation and population genomics data analysis. *Evolutionary Applications*, *11*(8), 1197-1211.

Hess, J. E., Matala, A. P., & Narum, S. R. (2011). Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources*, *11*, 137-149.

Hoarau, G., Boon, E., Jongma, D. N., Ferber, S., Palsson, J., Van der Veer, H. W., Rijnsdorp, A. D., Wytze, T. S., & Olsen, J. L. (2005). Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (Pleuronectes platessa L.). *Proceedings of the Royal Society of London B: Biological Sciences*, *272*(1562), 497-503.

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F ST. *Nature Reviews Genetics*, *10*(9), 639.

Hughes, A. R., Inouye, B. D., Johnson, M. T., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology letters*, *11*(6), 609-623.

Incze, L. S., Armstrong, D. A., & Wencker, D. L. (1982). Rates of development and growth of larvae of Chionoecetes bairdi and C. opilio in the southeastern Bering Sea. In *Proc. Int. Symp. Genus Chionoecetes. Fairbanks, AK: U. of Alaska Sea Grant College Program Rep* (Vol. 82, No. 10, pp. 191-218).

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, *11*(1), 94.

Jombart, T., & Collins, C. (2015). A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0. 0. *Imp Coll London-MRC Cent Outbreak Anal Model*, *43*.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Kidd, K. K., Pakstis, A. J., Speed, W. C., Lagacé, R., Chang, J., Wootton, S., Haigh, E., & Kidd, J. R. (2014). Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Science International: Genetics*, *12*, 215-224.

Knaus, B. J., & Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44-53.

Knutsen, H., Olsen, E. M., Jorde, P. E., Espeland, S. H., André, C., & Stenseth, N. C. (2011). Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. *Molecular ecology*, *20*(4), 768-783.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources*, *15*(5), 1179-1191.

Lal, M. M., Southgate, P. C., Jerry, D. R., & Zenger, K. R. (2016). Fishing for divergence in a sea of connectivity: the utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster Pinctada margaritifera. *Marine genomics*, *25*, 57-68.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS genetics*, *8*(1), e1002453.

Lemaire, C., Allegrucci, G., Naciri, M., Bahri-Sfar, L., Kara, H., & Bonhomme, F. (2000). Do discrepancies between microsatellite and allozyme variation reveal differential selection between sea and lagoon in the sea bass (Dicentrarchus labrax)?. *Molecular Ecology*, *9*(4), 457-467.

Lynch, M. (2008). Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. Molecular biology and evolution, 25(11), 2409-2419.

Nielsen, J. K., Taggart, S. J., Shirley, T. C., & Mondragon, J. (2007). Spatial distribution of juvenile and adult female Tanner crabs (Chionoecetes bairdi) in a glacial fjord ecosystem: implications for recruitment processes. *ICES Journal of Marine Science*, *64*(9), 1772-1784.

NPFMC The Plan Team for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands, K. Bush, M. Dorn, G. Eckert, R.J. Foy, B. Garber-Yonts, T. Hamazaki, J. N. Ianelli, A.E. Punt, L. Rugolo, M.S.M. Siddeek, W. Stockhausen, L. Slater, D. Stram, C. Szuwalski, B. J. Turnock, D. Weber and J. Zheng. (2016). "2016 Crab SAFE - North Pacific Fishery Management Council." http://www.npfmc.org/wp-content/PDFdocuments/resources/SAFE/CrabSAFE/2016CrabSAFE_final.pdf

Meyers, T. R., Botelho, C., Koeneman, T. M., & Short, S. (1990). Distribution of bitter crab dinoflagellate syndrome in southeast Alaskan Tanner crabs Chionoecetes. *Dis. Aquat. Org*, *9*, 37-43.

Merkouris, S. E., Seeb, L. W., & Murphy, M. C. (1998). Low levels of genetic diversity in highly exploited populations of Alaskan Tanner crabs, Chionoecetes bairdi, and Alaskan and Atlantic snow crabs, C. opilio. *Fishery Bulletin: National Oceanic and Atmospheric Administration, 96*, 525-537.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, *70*(12), 3321-3323.

Olsen, M. T., Pampoulie, C., Daníelsdóttir, A. K., Lidh, E., Bérubé, M., Víkingsson, G. A., & Palsbøll, P. J. (2014). Fin whale MDH-1 and MPI allozyme variation is not reflected in the corresponding DNA sequences. *Ecology and evolution*, *4*(10), 1787-1803.

Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological applications*, S146-S158.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, *2*(12), e190.

Paul, A. J. (1984). Mating frequency and viability of stored sperm in the Tanner crab Chionoecetes bairdi (Decapoda, Majidae). *Journal of Crustacean Biology*, *4*(3), 375-381.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS one, 7(5), e37135.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*(3), 608-627.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014). Demystifying the RAD fad. *Molecular ecology*, *23*(24), 5937-5942.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reitzel, A. M., Herrera, S., Layden, M. J., Martindale, M. Q., & Shank, T. M. (2013). Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular ecology*, *22*(11), 2953-2970.

Richar, J. I., Kruse, G. H., Curchitser, E., & Hermann, A. J. (2015). Patterns in connectivity and retention of simulated Tanner crab (Chionoecetes bairdi) larvae in the eastern Bering Sea. *Progress in Oceanography*, *138*, 475-485.

Rodríguez-Ezpeleta, N., Bradbury, I. R., Mendibil, I., Álvarez, P., Cotano, U., & Irigoien, X. (2016). Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: effects of sequence clustering parameters and hierarchical SNP selection. *Molecular ecology resources*.

Robertson, A. (1965). The interpretation of genotypic ratios in domestic animal populations. *Animal Science*, *7*(3), 319-324.

Sale, P. F., Cowen, R. K., Danilowicz, B. S., Jones, G. P., Kritzer, J. P., Lindeman, K. C., Planes, S., Polunin, N. V. C., Russ, G. R., Sadovy, Y. J., & Steneck, R. S. (2005). Critical science gaps impede use of no-take fishery reserves. *Trends in ecology & evolution*, *20*(2), 74-80.

Shanks, A. L. (2009). Pelagic larval duration and dispersal distance revisited. *The biological bulletin*, *216*(3), 373-385.

Sinnock, P. (1975). The Wahlund effect for the two-locus model. *The American Naturalist*, *109*(969), 565-570.

Somerton, D. A. (1980). A computer technique for estimating the size of sexual maturity in crabs. *Canadian Journal of Fisheries and Aquatic Sciences*, *37*(10), 1488-1494.

Somerton, D. A. (1981). Regional variation in the size of maturity of two species of tanner crab (Chionoecetes bairdi and C. opilio) in the eastern Bering Sea, and its use in defining management subareas. *Canadian Journal of Fisheries and Aquatic Sciences*, *38*(2), 163-174.

Spies, I., & Punt, A. E. (2015). The utility of genetics in marine fisheries management: a simulation study based on Pacific cod off Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(9), 1415-1432.

Spies, I., Spencer, P. D., & Punt, A. E. (2015). Where do we draw the line? A simulation approach for evaluating management of marine fish stocks with isolation-by-distance stock structure. *Canadian journal of fisheries and aquatic sciences*, *72*(7), 968-982.

Stevens, B. G., Haaga, J. A., & Donaldson, W. E. (1994). Aggregative mating of Tanner crabs, Chionoecetes bairdi. *Canadian Journal of Fisheries and Aquatic Sciences*, *51*(6), 1273-1280.

Stevens, B. G. (2003). Timing of aggregation and larval release by Tanner crabs, Chionoecetes bairdi, in relation to tidal current patterns. *Fisheries Research*, *65*(1), 201-216.

Stockhausen, W. T. (2018). "2018 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions" BSAI Crab SAFE https://www.npfmc.org/wp-content/PDFdocuments/resources/SAFE/CrabSAFE/2018/3-Tanner_SAFE_2018.pdf

Urban, D., & Hart, D. (1999). *Biology of the Tanner Crab, Chionoecetes Bairdi, in Alaska: A Report to the Alaska Board of Fisheries*. Alaska Department of Fish and Game, Division of Commercial Fisheries.

Van Wyngaarden, M., Snelgrove, P. V., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., Stanley, R. R. E. & Bradbury, I. R. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, Placopecten magellanicus, using RAD seq-derived SNP s. *Evolutionary applications*, *10*(1), 102-117.

Ward, R. D., Woodwark, M., & Skibinski, D. O. F. (1994). A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of fish biology*, *44*(2), 213-232.

Weersing, K., & Toonen, R. J. (2009). Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series*, *393*, 1-12.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *evolution*, *38*(6), 1358-1370.

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F ST. *The American Naturalist*, *186*(S1), S24-S36.

Wright, S. (1950). Genetical structure of populations. *Nature*, *166*, 247-49.

Xuereb, A., Benestan, L., Normandeau, E., Daigle, R. M., Curtis, J. M., Bernatchez, L., & Fortin, M. J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RAD seq, in a highly dispersive marine invertebrate (Parastichopus californicus). *Molecular ecology*, *27*(10), 2347-2364.

Zheng, J., & Kruse, G. H. (2006). Recruitment variation of eastern Bering Sea crabs: climate-forcing or top-down effects?. *Progress in Oceanography*, *68*(2-4), 184-204.

General Conclusions

The use of molecular data to address fisheries concerns is well established. For decades, population genetic estimates have been used to infer patterns of genetic exchange and determine independent fishery stocks (Carvalho & Hauser, 1995; Ovenden et al., 2015). The intent of this study was to use of high throughput sequencing techniques for resolving uncertainties about the genetic population structure of *C. bairdi* in Alaskan waters. This was the first study since the 1990's to measure genetic variability of *C. bairdi* in Alaska, and it comprises the largest dataset of this type that has been collected for an Alaskan crustacean species. Using a double digest reduced-representation sequencing (ddRADseq) methodology, we acquired an average of 1.9 M sequence reads per individual, which aligned to 21,211 loci across the genome. The results of this study, based on 2,740 neutral SNP markers, do not support population subdivisions. Estimates of genetic diversity between sampling regions were not significantly different from zero, the null hypothesis of panmixia.

These results do not meet the same conclusions as previous genetic population structure studies on this species. Allozymes are significantly differentiated between Southeast Alaska, the Gulf of Alaska, and the Bering Sea, and are differentiated east and west of 162° W Longitude in the Bering Sea (Merkouris et al., 1998). However, differences among allozymes could be associated with phenotypic plasticity or selection rather than genetic drift associated with subpopulation isolation. Other studies using nuclear markers have failed to detect significant neutral genetic variation between regions that have allozyme divergence (Lemaire et al., 2000; Olsen et al., 2014). Variation in mitochondrial haplotypes supports the idea that *C. bairdi* larvae are transported by the Alaska Coastal Current (ACC), but the evidence is not statistically robust. No independent nucleotide variants are unique to a single region, but haplotypes do indicate differentiation between regions. The patterns of haplotype diversity in regions downstream of the ACC could suggest gene flow from East to West, from Southeast Alaska to the Bering Sea (Bunch et al., 1998). Our results of panmixia also suggest gene flow, though not necessarily directional.

To put our results in the context of other genetic studies on Alaskan crab species, Red King Crab (*Paralithodes camtschaticus*) are affected by barriers to gene flow across the state. Population

genetic structure of *P. camtschaticus* is marked by isolated populations in Southeast AK, and one genetically homogenous population throughout the Eastern Bering Sea and Western Gulf of Alaska (Grant & Cheng, 2012). *Chionoecetes opilio* have strong genetic evidence of panmictic population genetics (Albrecht et al., 2014). Given that *C. bairdi* and *C. opilio* occupy overlapping ranges and share enough life history traits that they are able to hybridize, it is plausible that they also demonstrate a similar lack of genetic differentiation between regional areas either because of larval dispersal or adult migration patterns.

After all of the SNP filtering steps in this study to remove low frequency alleles, only one individual from the Eastern Bering Sea region west of 166° W Longitude looked like a potential *Chionoecetes* hybrid based on PCA. That individual was retained in the final dataset because we were not able to verify that it was a hybrid or backcrossed individual. I did collect an additional smaller ddRADseq dataset, which was not reported in the chapter. That dataset contained both *C. bairdi* and *C. opilio* samples, and the two species were distinguishable based on 3850 SNPs. The southern extent of the *C. opilio* range in the Bering Sea has shifted northward in periods of warm years (Orensanz et al., 2005). The extent of hybridization in the future could change based on these range shifts, and species management may benefit from a large scale RADseq study of both species. From an evolutionary standpoint, it may be possible for temperature related variants from either species to be exchanged and improve overall fitness under changing climate scenarios.

Reduced-representation genome sequencing methods are gaining traction in the field of population genetics and are increasingly applied to study the genetic population structure of non-model organisms (Puritz et al., 2014). The ability to detect a greater quantity of variable sites through high throughput sequencing approaches to resolve questions of fine-scale genetic population structure is enticing. Yet, as evidenced in this study, large datasets of SNP genotypes will not necessarily provide the resolution necessary to assign individuals back to their regions of origin, because the genetic structure does not exist in the natural population. We could not identify sufficient variation that was unique to management regions (Benestan et al., 2015; Knutsen et al., 2011) so our assignment tests were not better than random assignment. While multilocus genotype sets provide sufficient assignment power when applied to other marine

organisms, the data we collected were not sufficient to distinguish between major geographic regions.

The average estimates of genetic diversity that we observed fell within ranges observed in other RADseq studies of marine invertebrate species (Reitzel et al., 2013; Xuereb et al., 2018). One notable difference was that we consistently observed average locus heterozygosity within regions that was lower than expected for populations in Hardy-Weinberg equilibrium. So *C. bairdi* might have higher rates on inbreeding compared to other marine invertebrates. The ability of *C. bairdi* females to retain sperm from previous mating events and fertilize later clutches do violate our assumptions of non-overlapping generations and random mating (Christiansen, 1988). This could be why *C. bairdi* appear to be inbred by these estimates while other species with pelagic larval stages and variability in settlement survival do not. We did not detect significant genetic differentiation between sampling regions for *C. bairdi*, but other studies on marine species using RADseq datasets of this size have been used to detect population structure on finer geographic scales (Benestan et al., 2015; Xuereb et al., 2018).

We chose to study *C. bairdi* because of its local importance as a harvested species, and because this species has management concerns across the state that can benefit from information on the genetic population structure. The concern that PWS stocks are influenced by non-local recruitment is not disproven by our results (Berceli et al., 2002). The samples from the PWS region were not significantly genetically differentiated from any of the other regions in this study, so there could be influx to this area. As mentioned earlier, the distinction between the eastern and western sides of 166°W Longitude is contentious. We were unable to verify its relevance based genetic differentiation estimates. The idea of an isolated self-propagating subpopulation in the Bering Sea east of 166°W is not supported by these data and standard population genetic analyses.

If there is further interest in understanding the nature of connectivity between these regions, the data collected from this project may still be applicable. Toolsets for utilizing the full sequence of loci, to identify SNPs that are linked, termed microhaplotypes, rather than individual independent SNPs are under development (Catchen et al., 2013; Hendricks et al., 2018).

Microhaplotypes could potentially encode more information about the evolutionary histories and relatedness of individuals in this dataset, and might therefore provide a better estimate of multigenerational patterns of connectivity along geographical scales (Baetscher et al., 2018; Kidd et al., 2014; Lawson et al., 2012).

References

Albrecht, G. T., Valentin, A. E., Hundertmark, K. J., & Hardy, S. M. (2014). Panmixia in Alaskan populations of the snow crab Chionoecetes opilio (Malacostraca: Decapoda) in the Bering, Chukchi, and Beaufort Seas. *Journal of Crustacean Biology*, *34*(1), 31-39

Baetscher, D. S., Clemento, A. J., Ng, T. C., Anderson, E. C., & Garza, J. C. (2018). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources*, *18*(2), 296-305.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). *Molecular ecology*, *24*(13), 3299-3315.

Berceli, R., Trowbridge, C., Bechtol, W. R., Rue–Commissioner, F., & Mecum–Director, R. D. (2002). Review of the King and Tanner crab fisheries in Prince William Sound. Regional Information Report No. 2A02-06.

Bunch, T., Highsmith, R. C., & Shields, G. F. (1998). Genetic evidence for dispersal of larvae of Tanner crabs (Chionoecetes bairdi) by the Alaskan Coastal Current. *Molecular Marine Biology and Biotechnology*, 7(2), 153-159.

Carvalho, G. R., & Hauser, L. (1995). Molecular genetics and the stock concept in fisheries. In *Molecular genetics in fisheries* (pp. 55-79). Springer, Dordrecht.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, *22*(11), 3124-3140.

Christiansen, F. B. (1988). The Wahlund effect with overlapping generations. *The American Naturalist*, *131*(1), 149-156.

Grant, W. S., & Cheng, W. (2012). Incorporating deep and shallow components of genetic structure into the management of Alaskan red king crab. *Evolutionary applications*, *5*(8), 820-837.

Hendricks, S., Anderson, E. C., Antao, T., Bernatchez, L., Forester, B. R., Garner, B., & Sethuraman, A. (2018). Recent advances in conservation and population genomics data analysis. *Evolutionary Applications*, *11*(8), 1197-1211.

Kidd, K. K., Pakstis, A. J., Speed, W. C., Lagacé, R., Chang, J., Wootton, S., & Kidd, J. R. (2014). Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Science International: Genetics*, *12*, 215-224.

Knutsen, H., Olsen, E. M., Jorde, P. E., Espeland, S. H., André, C., & Stenseth, N. C. (2011). Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. *Molecular ecology*, *20*(4), 768-783.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS genetics*, *8*(1), e1002453.

Lemaire, C., Allegrucci, G., Naciri, M., Bahri-Sfar, L., Kara, H., & Bonhomme, F. (2000). Do discrepancies between microsatellite and allozyme variation reveal differential selection between sea and lagoon in the sea bass (Dicentrarchus labrax)?. *Molecular Ecology*, *9*(4), 457-467.

Merkouris, S. E., Seeb, L. W., & Murphy, M. C. (1998). Low levels of genetic diversity in highly exploited populations of Alaskan Tanner crabs, Chionoecetes bairdi, and Alaskan and Atlantic snow crabs, C. opilio. *Fishery Bulletin: National Oceanic and Atmospheric Administration, 96*, 525-537.

Olsen, M. T., Pampoulie, C., Daníelsdóttir, A. K., Lidh, E., Bérubé, M., Víkingsson, G. A., & Palsbøll, P. J. (2014). Fin whale MDH-1 and MPI allozyme variation is not reflected in the corresponding DNA sequences. *Ecology and evolution*, *4*(10), 1787-1803.

Orensanz, J. M., Ernest, B., Armstrong, D. A., Stabeno, P., & Livingston, P. (2005). Contraction of the geographic range of distribution of snow crab (Chionoecetes opilio) in the eastern Bering Sea: an environmental ratchet?. *Scripps Institution of Oceanography; California Cooperative Oceanic Fisheries Investigations Reports*; 45; 8-2005; 65-79.

Ovenden, J. R., Berry, O., Welch, D. J., Buckworth, R. C., & Dichmont, C. M. (2015). Ocean's eleven: a critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. *Fish and Fisheries*, *16*(1), 125-159.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014). Demystifying the RAD fad. *Molecular ecology*, *23*(24), 5937-5942.

Reitzel, A. M., Herrera, S., Layden, M. J., Martindale, M. Q., & Shank, T. M. (2013). Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular ecology*, *22*(11), 2953-2970.

Xuereb, A., Benestan, L., Normandeau, E., Daigle, R. M., Curtis, J. M., Bernatchez, L., & Fortin, M. J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RAD seq, in a highly dispersive marine invertebrate (Parastichopus californicus). *Molecular ecology*, *27*(10), 2347-2364.

Appendix A

Appendix A: Collection and sequence filtering information for each individual sample.

Individual sequence ID, sampling region, Latitude and Longitude coordinates in decimal degrees, the year of sample collection, sequencing provider, number of raw sequence reads , number of reads remaining after first quality filters, the number of aligned clusters identified within the individual, number of clusters that passed clustering depth thresholds ($10 < x < 40{,}000$), heterozygosity estimates, error of the H estimate, the number of clusters that were identified as consensus reads, number of loci from each individual that are included in the initial genotype assembly.

Individual ID	Region	Latitude (deg)	Longitude (deg)	Collection Year	Sex	Sequencing Provider	reads raw	reads passed filter	clusters total	clusters high depth	heterozygosity estimate	error estimate	reads consensus	loci in assembly
4704-plate1-A01-AL2016-939-MiSeqNEBUltra_S69_L000001	EBS_NW	-	-	2012	-	RTL	815170	808878	43703	10348	0.009624	0.005025	9930	4470
4704-plate1-A02-AL2016-481-MiSeqNEBUltra_S77_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	1733434	81733	13995	0.010128	0.004352	13429	5904
4704-plate1-A03-AL2016-940-MiSeqNEBUltra_S85_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	1562949	67426	13329	0.010139	0.004197	12804	5711
4704-plate1-B01-AL2016-653-MiSeqNEBUltra_S70_L000001	EBS_NW	-	-	2012	-	RTL	3E+06	2722029	93817	16196	0.011177	0.003473	15508	6915
4704-plate1-B02-AL2016-482-MiSeqNEBUltra_S78_L000001	EBS_NW	-	-	2012	-	RTL	828998	822315	38720	9886	0.010377	0.004522	9485	4234
4704-plate1-B03-AL2016-885-MiSeqNEBUltra_S86_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	2072982	78178	14763	0.010236	0.003594	14188	6360
4704-plate1-C01-AL2016-654-MiSeqNEBUltra_S71_L000001	EBS_NW	-	-	2012	-	RTL	1E+06	1196108	52306	12188	0.009968	0.004003	11705	5254
4704-plate1-C02-AL2016-876-MiSeqNEBUltra_S79_L000001	EBS_NW	-	-	2012	-	RTL	1E+06	1464865	52216	12356	0.0102	0.003879	11864	5336
4704-plate1-C03-AL2016-941-MiSeqNEBUltra_S87_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	2360614	79554	15370	0.010883	0.003587	14725	6559
4704-plate1-D01-AL2016-655-MiSeqNEBUltra_S72_L000001	EBS_NW	-	-	2012	-	RTL	1E+06	1205349	53890	12273	0.010267	0.005499	11824	5259
4704-plate1-D03-AL2016-943-MiSeqNEBUltra_S88_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	2332930	77796	15014	0.010045	0.00525	14388	6337
4704-plate1-E01-AL2016-656-MiSeqNEBUltra_S73_L000001	EBS_NW	-	-	2012	-	RTL	1E+06	1170107	47463	12317	0.009767	0.004198	11866	5298
4704-plate1-E02-AL2016-878-MiSeqNEBUltra_S81_L000001	EBS_NW	-	-	2012	-	RTL	1E+06	1395691	64076	13480	0.011898	0.004406	12938	5104
4704-plate1-E03-AL2016-944-MiSeqNEBUltra_S89_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	1777987	71583	14082	0.010438	0.003895	13502	5997
4704-plate1-F01-AL2016-662-MiSeqNEBUltra_S74_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	2048537	67999	14778	0.010715	0.003557	14162	6373
4704-plate1-F02-AL2016-882-MiSeqNEBUltra_S82_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	1616886	66857	13751	0.009931	0.004011	13194	5883
4704-plate1-F03-AL2016-945-MiSeqNEBUltra_S90_L000001	EBS_NW	-	-	2012	-	RTL	849988	842050	45147	10458	0.010275	0.004335	10012	4457
4704-plate1-G01-AL2016-479-MiSeqNEBUltra_S75_L000001	EBS_NW	-	-	2012	-	RTL	1E+06	1476095	55287	13254	0.010149	0.004019	12712	5706
4704-plate1-G02-AL2016-883-MiSeqNEBUltra_S83_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	1632785	87109	13463	0.009922	0.004933	12905	5658
4704-plate1-G03-AL2016-946-MiSeqNEBUltra_S91_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	1649957	59003	12942	0.010241	0.004047	12403	5567
4704-plate1-H01-AL2016-480-MiSeqNEBUltra_S76_L000001	EBS_NW	-	-	2012	-	RTL	692869	687441	40957	9663	0.010027	0.004555	9261	4169
4704-plate1-H02-AL2016-884-MiSeqNEBUltra_S84_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	1585511	59310	13273	0.010223	0.003753	12738	5672
4704-plate1-H03-AL2016-658-MiSeqNEBUltra_S92_L000001	EBS_NW	-	-	2012	-	RTL	2E+06	1970675	61754	14133	0.010032	0.003497	13547	6083
4704-plate2-A01-AL2016-872-MiSeqNEBUltra_S101_L000001	EBS_SE	-	-	2012	-	RTL	3E+06	2668858	93746	15706	0.010694	0.004335	15028	6638
4704-plate2-A02-AL2016-477-MiSeqNEBUltra_S109_L000001	EBS_SE	-	-	2012	-	RTL	3E+06	3005824	148485	16483	0.011189	0.003806	15749	6758
4704-plate2-A03-AL2016-928-MiSeqNEBUltra_S117_L000001	EBS_SE	-	-	2012	-	RTL	855603	846143	47747	10835	0.009745	0.005222	10418	4675
4704-plate2-A04-AL2016-937-MiSeqNEBUltra_S124_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	1815538	73752	13648	0.010051	0.003743	13129	5842
4704-plate2-B01-AL2016-648-MiSeqNEBUltra_S102_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	1807623	83219	13789	0.010487	0.003798	13208	5886
4704-plate2-B02-AL2016-866-MiSeqNEBUltra_S110_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	2008416	118753	14635	0.010897	0.003711	14016	6141
4704-plate2-B03-AL2016-929-MiSeqNEBUltra_S118_L000001	EBS_SE	-	-	2012	-	RTL	1E+06	1346743	60938	13068	0.00987	0.004206	12549	5651
4704-plate2-C01-AL2016-650-MiSeqNEBUltra_S103_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	2386256	79694	15283	0.010993	0.00368	14623	6520
4704-plate2-C02-AL2016-867-MiSeqNEBUltra_S111_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	2227982	83946	15188	0.01095	0.003681	14523	6475
4704-plate2-C03-AL2016-930-MiSeqNEBUltra_S119_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	2054770	75092	14719	0.010629	0.003783	14095	6322
4704-plate2-D01-AL2016-472-MiSeqNEBUltra_S104_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	2113364	85711	14678	0.010268	0.005682	14029	6131
4704-plate2-D02-AL2016-868-MiSeqNEBUltra_S112_L000001	EBS_SE	-	-	2012	-	RTL	3E+06	2605741	140218	15898	0.011539	0.005217	15162	6552
4704-plate2-E01-AL2016-473-MiSeqNEBUltra_S105_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	1766658	65007	14076	0.010548	0.003873	13459	5997
4704-plate2-E02-AL2016-870-MiSeqNEBUltra_S113_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	1912097	85392	14479	0.010751	0.00386	13860	6184
4704-plate2-E03-AL2016-932-MiSeqNEBUltra_S120_L000001	EBS_SE	-	-	2012	-	RTL	1E+06	1158005	58072	12463	0.00998	0.006225	11963	5303
4704-plate2-F01-AL2016-474-MiSeqNEBUltra_S106_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	1590571	70486	13130	0.010172	0.00399	12609	5591
4704-plate2-F02-AL2016-871-MiSeqNEBUltra_S114_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	2272639	132523	15244	0.010963	0.003786	14578	6351
4704-plate2-F03-AL2016-933-MiSeqNEBUltra_S121_L000001	EBS_SE	-	-	2012	-	RTL	1E+06	1235956	51888	12534	0.009983	0.004314	12028	5436
4704-plate2-G01-AL2016-475-MiSeqNEBUltra_S107_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	1616705	62163	13733	0.009951	0.004223	13139	5887
4704-plate2-G02-AL2016-873-MiSeqNEBUltra_S115_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	1649700	64044	13620	0.01024	0.004177	13070	5821
4704-plate2-G03-AL2016-934-MiSeqNEBUltra_S122_L000001	EBS_SE	-	-	2012	-	RTL	4E+06	3723846	118540	18077	0.011863	0.003486	17272	7563
4704-plate2-H01-AL2016-476-MiSeqNEBUltra_S108_L000001	EBS_SE	-	-	2012	-	RTL	1E+06	1026190	82064	11127	0.009983	0.004314	10646	4672
4704-plate2-H02-AL2016-874-MiSeqNEBUltra_S116_L000001	EBS_SE	-	-	2012	-	RTL	2E+06	1777045	72287	13783	0.010252	0.003587	13231	5860
4704-plate2-H03-AL2016-936-MiSeqNEBUltra_S123_L000001	EBS_SE	-	-	2012	-	RTL	1E+06	1153876	52211	12113	0.009595	0.004543	11625	5259
4704-plate3-A01-AL2016-673-MiSeqNEBUltra_S132_L000001	PWS	60.594	-146.21883	2009	M	RTL	612998	605818	50891	8579	0.009868	0.005446	8189	3654
4704-plate3-A02-AL2016-952-MiSeqNEBUltra_S140_L000001	PWS	60.556	-146.46517	2009	F	RTL	2E+06	1585602	76012	13310	0.010057	0.004539	12784	5623
4704-plate3-B02-AL2016-955-MiSeqNEBUltra_S141_L000001	PWS	60.56	-146.50383	2009	M	RTL	1E+06	1279967	63470	12312	0.01019	0.00415	11804	5230
4704-plate3-B03-AL2016-678-MiSeqNEBUltra_S149_L000001	PWS	60.556	-146.46517	2009	F	RTL	2E+06	1864790	105025	14243	0.010486	0.003784	13650	6074
4704-plate3-C01-AL2016-675-MiSeqNEBUltra_S134_L000001	PWS	60.559	-146.1465	2009	F	RTL	926330	912851	64501	10989	0.010215	0.004519	10517	4704
4704-plate3-C02-AL2016-226-MiSeqNEBUltra_S142_L000001	PWS	-	-	2009	M	RTL	2E+06	1656739	90613	13809	0.010566	0.00392	13215	5834
4704-plate3-C03-AL2016-679-MiSeqNEBUltra_S150_L000001	PWS	60.556	-146.46517	2009	F	RTL	2E+06	2431637	133555	15893	0.011475	0.003752	15146	6658
4704-plate3-D01-AL2016-683-MiSeqNEBUltra_S135_L000001	PWS	60.504	-146.435	2009	F	RTL	1E+06	1104417	79120	11778	0.009989	0.005665	11301	4994
4704-plate3-D02-AL2016-230-MiSeqNEBUltra_S143_L000001	PWS	-	-	2009	M	RTL	1E+06	1104850	78753	11798	0.010181	0.005718	11286	4994

4704-plate3-D03-AL2016-680-MiSeqNEBUltra_S151_L000001	PWS	60.556	-146.46517	2009	F	RTL	2E+06	1891571	104298	14439	0.010066	0.005346	13800	5925
4704-plate3-E01-AL2016-685-MiSeqNEBUltra_S136_L000001	PWS	60.56	-146.50383	2009	M	RTL	1E+06	1132534	74844	11797	0.010528	0.004333	11297	4963
4704-plate3-E02-AL2016-463-MiSeqNEBUltra_S144_L000001	PWS	60.594	-146.21883	2009	F	RTL	1E+06	1245695	83574	11968	0.010367	0.004327	11473	4998
4704-plate3-E03-AL2016-750-MiSeqNEBUltra_S152_L000001	PWS	60.556	-146.46517	2009	M	RTL	2E+06	2140543	120509	14767	0.010547	0.003707	14153	6254
4704-plate3-F01-AL2016-687-MiSeqNEBUltra_S137_L000001	PWS	60.71	-146.16933	2009	F	RTL	916511	907027	56714	10802	0.010226	0.004531	10335	4579
4704-plate3-F02-AL2016-465-MiSeqNEBUltra_S145_L000001	PWS	60.594	-146.21883	2009	F	RTL	1E+06	1330706	77885	12278	0.010276	0.004152	11765	5144
4704-plate3-F03-AL2016-751-MiSeqNEBUltra_S153_L000001	PWS	60.556	-146.46517	2009	M	RTL	3E+06	3292411	120228	17151	0.011177	0.003518	16361	7272
4704-plate3-G01-AL2016-748-MiSeqNEBUltra_S138_L000001	PWS	60.559	-146.1465	2009	M	RTL	2E+06	1491168	100434	12666	0.010295	0.004065	12128	5271
4704-plate3-G02-AL2016-466-MiSeqNEBUltra_S146_L000001	PWS	60.559	-146.1465	2009	M	RTL	1061165	1049994	69584	11557	0.010177	0.004363	11088	4897
4704-plate3-G03-AL2016-752-MiSeqNEBUltra_S154_L000001	PWS	60.556	-146.46517	2009	M	RTL	2570458	2544261	106220	15596	0.01072	0.003848	14924	6669
4704-plate3-H01-AL2016-757-MiSeqNEBUltra_S139_L000001	PWS	60.71	-146.16933	2009	F	RTL	1366241	1354581	87249	12332	0.010006	0.003923	11845	5153
4704-plate3-H02-AL2016-467-MiSeqNEBUltra_S147_L000001	PWS	60.559	-146.1465	2009	M	RTL	1072476	1063678	56151	11337	0.009995	0.004075	10881	4851
4704-plate3-H03-AL2016-755-MiSeqNEBUltra_S155_L000001	PWS	60.504	-146.435	2009	M	RTL	2670198	2650419	98476	15010	0.010771	0.003448	14368	6367
4704-plate4-A01-AL2016-886-MiSeqNEBUltra_S164_L000001	SEAK	58.3354639	-134.71854	2016	M	RTL	2572363	2549682	139593	15821	0.010914	0.00422	15117	6604
4704-plate4-A02-AL2016-894-MiSeqNEBUltra_S172_L000001	SEAK	58.3191694	-134.86906	2016	M	RTL	2220212	2201125	92602	14997	0.01055	0.003771	14358	6346
4704-plate4-A03-AL2016-907-MiSeqNEBUltra_S180_L000001	SEAK	58.185825	-134.35015	2016	F	RTL	1692041	1679763	79365	13812	0.010145	0.004644	13227	5828
4704-plate4-B01-AL2016-887-MiSeqNEBUltra_S165_L000001	SEAK	58.3349889	-134.7171	2016	F	RTL	2694679	2672771	133361	16545	0.012111	0.003814	15823	7021
4704-plate4-B02-AL2016-896-MiSeqNEBUltra_S173_L000001	SEAK	58.3188278	-134.88407	2016	F	RTL	2239336	2220602	122455	14986	0.011115	0.004022	14341	6239
4704-plate4-B03-AL2016-908-MiSeqNEBUltra_S181_L000001	SEAK	58.1839278	-134.35026	2016	M	RTL	2949812	2930007	119460	16413	0.011039	0.003707	15678	6978
4704-plate4-C01-AL2016-888-MiSeqNEBUltra_S166_L000001	SEAK	58.3522861	-134.70116	2016	M	RTL	1178198	1168364	57549	10734	0.010301	0.004157	10299	4499
4704-plate4-C02-AL2016-897-MiSeqNEBUltra_S174_L000001	SEAK	58.1853694	-134.60178	2016	F	RTL	2990201	2966006	93780	15929	0.010963	0.0035	15191	6680
4704-plate4-C03-AL2016-909-MiSeqNEBUltra_S182_L000001	SEAK	58.4513917	-134.81934	2016	M	RTL	4827910	4793678	293544	70605	0.00807	0.00444	67024	19587
4704-plate4-D01-AL2016-889-MiSeqNEBUltra_S167_L000001	SEAK	58.3676361	-134.68379	2016	F	RTL	2006793	1987098	195804	41669	0.008702	0.006867	39192	17219
4704-plate4-D02-AL2016-898-MiSeqNEBUltra_S175_L000001	SEAK	58.2014167	-134.41874	2016	M	RTL	2413532	2393008	145686	15189	0.010678	0.005091	14513	6265
4704-plate4-D03-AL2016-910-MiSeqNEBUltra_S183_L000001	SEAK	58.4511389	-134.78352	2016	M	RTL	3047830	3022060	313800	41831	0.009419	0.006171	39365	19509
4704-plate4-E01-AL2016-890-MiSeqNEBUltra_S168_L000001	SEAK	58.3520361	-134.66727	2016	F	RTL	658306	650257	162760	6308	0.017642	0.007449	5661	2274
4704-plate4-E02-AL2016-901-MiSeqNEBUltra_S176_L000001	SEAK	58.2026639	-134.41867	2016	M	RTL	3088921	3066760	157203	16714	0.010964	0.003518	15988	6911
4704-plate4-E03-AL2016-911-MiSeqNEBUltra_S184_L000001	SEAK	58.4188306	-134.81777	2016	F	RTL	2738265	2720001	92256	15917	0.010829	0.003608	15252	6720
4704-plate4-F01-AL2016-891-MiSeqNEBUltra_S169_L000001	SEAK	58.3341944	-134.66687	2016	M	RTL	3124733	3100035	123863	16671	0.01115	0.003706	15917	7016
4704-plate4-F02-AL2016-903-MiSeqNEBUltra_S177_L000001	SEAK	58.1858611	-134.35269	2016	M	RTL	3010723	2988994	99213	16101	0.011022	0.003588	15379	6793
4704-plate4-F03-AL2016-912-MiSeqNEBUltra_S185_L000001	SEAK	58.4186722	-134.78495	2016	M	RTL	3682665	3655698	310977	63424	0.007947	0.004833	60092	17632
4704-plate4-G01-AL2016-892-MiSeqNEBUltra_S170_L000001	SEAK	58.3183667	-134.8685	2016	M	RTL	2160588	2143340	137118	14744	0.010637	0.004433	14098	6096
4704-plate4-G02-AL2016-904-MiSeqNEBUltra_S178_L000001	SEAK	58.1858611	-134.35269	2016	M	RTL	2024661	2009114	73045	14222	0.010407	0.003899	13628	6069
4704-plate4-G03-AL2016-913-MiSeqNEBUltra_S186_L000001	SEAK	58.4022611	-134.78449	2016	M	RTL	2408076	2391798	191009	15925	0.013717	0.004116	14853	6426
4704-plate4-H01-AL2016-893-MiSeqNEBUltra_S171_L000001	SEAK	58.318625	-134.86928	2016	M	RTL	2713422	2694186	131561	15371	0.013029	0.003425	14695	6346
4704-plate4-H02-AL2016-905-MiSeqNEBUltra_S179_L000001	SEAK	58.185825	-134.35015	2016	M	RTL	2441045	2424363	80285	15015	0.010282	0.003487	14406	6429
4704-plate4-H03-AL2016-915-MiSeqNEBUltra_S187_L000001	SEAK	58.401875	-134.76875	2016	M	RTL	1519300	1508743	108386	12966	0.01123	0.004075	12380	5347